UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

January
Revised

SIPP User Note Subjects

Rectangular Files: 1984 Panel

Wave 1

No. 1 Universe Description Correction

No. 2 Revised Documentation Pages

No. 3 Changes in Machine-Readable Dictionary

No. 4 Imputation Explanation

No. 5 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty."

Wave 2

No. 1 Matching Person Records Between Wave 1 and Wave 2 Public Use Files

No. 2 Matching the Wave 1 and Wave 2 SIPP Data Files

No. 3 Comparison of Item Locations on Wave 1 and Later Waves

No. 4 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty."

No. 5 MSA Coding Correction

Wave 3

No. 1 Comparison of Wave 3 Data Dictionary to Wave 2 Data Dictionary

No. 2 Revised Data Dictionary Pages

No. 3 Matching the 1984 SIPP Panel, Wave 1, Rotation 4 and Wave 3, Rotation 4

No. 4 Reasons for Mismatches between Wave 2 and Wave 3

No. 5 Revised Data Dictionary Pages (pages 117-118)

No. 6 Comparison of Item Locations on Wave 1 and Later Waves

No. 7 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty"

No. 8 Revised Documentation Page (page 2)

Wave 3 (Core and Topical Module)

No. 1 Comparison of Wave 3 Data Dictionary to Wave 2 Data Dictionary

No. 2 Revised Data Dictionary pages

No. 3 Matching the 1984 SIPP Panel, Wave 1 Rotation 4 and Wave 3 Rotation 4

No. 4 Reasons for Mismatches between Wave 2 and Wave 3

No. 5 Revised Data Dictionary Pages (pages 117-118)

No. 6 Comparison of Item Locations on Wave 1 and Later Waves

No. 7 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty"

No. 8 Revised Documentation Pages (pages 493, 495,537)

No. 9 Revised Documentation Page (page 2)

No. 10 Specific Imputation Flags in the Topical Module

No. 11 Correction to Topical Module - Disability Status of Children

Wave 4

No. 1 Resolution of Nonmatches: 1984 SIPP Panel, Waves 3 and 4 and Armed Forces Status Incorrectly Classified as "Activ Duty"

No. 2 Sizes of AFDC Amount Fields

Wave 4 (Core and Topical Module)

No. 1 Resolution of Nonmatches: 1984 SIPP Panel, Waves 3 and 4 and Armed Forces Status Incorrectly Classified as "Active
Duty"

No. 2 Revised Documentation Page (page 562)

No. 3 Variable Changes on Second File Release

No. 4 Sizes of AFDC Amount Fields

No. 5 Revised Documentation Page (page 21)

Wave 5 (Core)

No. 1 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty" (Relationship to this Wave)

No. 2 Matching Information for Waves 5 and 6

No. 3 Resolution of Nonmatches; 1984 SIPP Panel, Waves 4 and 5

No. 4 Sizes of AFDC Amount Fields

No. 5 Revised Documentation Page (page 21)

No. 6 Revised Documentation Page (page 21)

Wave 5 Core and Topical Module

No. 1 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty" (Relationship to this Wave)

No. 2 Matching Information for Waves 5 and 6

No. 3 Resolution of Nonmatches; 1984 SIPP Panel, Waves 4 and 5

No. 4 Sizes of AFDC Amount Fields

Wave 6 (Core)

No. 1 Matching Information for Waves 5 and 6

No. 2 Sizes of AFDC Amount Fields

No. 3 Revised Documentation Page (page 21)

Wave 7 (Core)

No. 1 Revised Documentation Page (page 21) (12/86)

No. 2 Sizes of AFDC Amount Fields

No. 3 Revised Documentation Page (page 21) (5/87)

Wave 7 (Core and Topical Module)

No. 1 "A" and "B" Parameters

No. 2 Pension Plan Parameters

No. 3 Data File/Data Dictionary Inconsistency

No. 4 Revised Documentation Pages (183-224)

Wave 8 (Core)

No. 1 Revised Documentation Page (page 21)(12/86)

No. 2 Sizes of AFDC Amount Fields

No. 3 Revised Documentation Page (page 21)(5/87)

Wave 8 (Core and Topical Module)

No. 1 Reference Person - Field TM8268

No. 2 Fertility Module

Wave 9 (Core)

No. 1 Revised Documentation Page (page 21)

No. 2 Explanation for Additional Family Records in Version 2

Rectangular Files: 1985 Panel

Wave 1

No user notes issued as of 6/89

Wave 2

No. 1 Correction of Specific Mnemonics with Duplicate Names

No. 2 New Descriptions for SC0064, SC0066, and SC0068

No. 3 Omission of Mnemonic SE1-OCC

Wave 3

No. 1 Correction of Specific Mnemonics with Duplicate Names

No. 2 New Descriptions for SC0064, SC0066, and SC0068

No. 3 Evaluations of Topical Module Data

Wave 4

No. 1 Correction of Specific Mnemonics with Duplicate Names

No. 2 New Descriptions for SC0064, SC0066, and SC0068

Wave 6

No user notes issued as of 6/89.

Relational Files: 1984 Panel

Wave 1

No. 1 Revised Documentation Page (4-24B and 4-24C)

No. 2 File Changes in Second Release of File

No. 3 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty"

Wave 2

No. 1 Matching Person Records Between Wave 1 and Wave 2 Public Use Files

No. 2 Revised Version of the Machine-readable Data Dictionary and Location of Variable AFDC3

No. 3 Matching the Wave 1 and Wave 2 SIPP Data Files

No. 4 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty."

No. 5 Comparison of Item Locations on Wave 1 and Later Waves

No. 6 MSA Coding Correction

Wave 3

No. 1 Comparison of Wave 3 Data Dictionary to Wave 2 Data Dictionary

No. 2 Revised Data Dictionary pages

No. 3 Matching the 1984 SIPP Panel, Wave 1 Rotation 4 and Wave 3 Rotation 4

No. 4 Matching the 1984 SIPP Panel, Waves 2 and 3, Rotations 1, 2, and 3

No. 5 Weighting on Revised File

No. 6 Comparison of Item Locations on Wave 1 and Later Waves

No. 7 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status was Incorrectly Classified as "Active Duty"

No. 8 MSA Coding Correction

Wave 4

No. 1 Resolution of Nonmatches: 1984 SIPP Panel, Waves 3 and 4 and Armed Forces Status Incorrectly Classified as "Activ Duty"

No. 2 MSA Coding Correction

Wave 5

No. 1 SIPP 1984 Panel Wave 1 and Wave 2 Cases where Armed Forces Status Was Incorrectly Classified as "Active Duty" (Relationship to this Wave)

No. 2 MSA Coding Correction

No. 3 Matching Information for Waves 5 and 6

No. 4 Resolution of Nonmatches; 1984 SIPP Panel, Waves 4 and 5

No. 5 Revised Documentation Page (page 2)

No. 6 Revised Documentation Page (page 21)

Wave 6

No. 1 Matching Information for Waves 5 and 6

No. 2 Revised Documentation Page (page 2)

No. 3 Revised Documentation Page (page 21)

Wave 7

No. 1 Revised Documentation Page (page 21) (12/86)

No. 2 Revised Documentation Page (page 323)

No. 3 Revised Documentation Page (page 21) (5/87)

Wave 8

No. 1 Revised Documentation Page (page 21) (12/86)

No. 2 Revised Documentation Page (page 323)

No. 3 Revised Documentation Page (page 21) (5/87)

Wave 9

No. 1 Revised Documentation Page (page 21)

Relational Files: 1985 Panel

Wave 1

No user notes issued as of 6/89

Wave 2

No. 1 Correction of Specific Mnemonics with Duplicate Names
No. 2 New Descriptions for SC064, SC0066, and SC0068

Waves 3-8

No user notes issued as of 6/89

SURVEY OF INCOME AND PROGRAM PARTICIPATION USERS' GUIDE

Washington, D.C.

July 1987

U.S. DEPARTMENT OF COMMERCE

Clarence J. Brown, Acting Secretary
Robert Ortner, Acting Under Secretary
for Economic Affairs

BUREAU OF THE CENSUS

John G. Keane, Director
C. Louis Kincannon, Deputy Director

PREFACE

This Users' Guide is intended as a reference to which the user can turn for understanding of the Survey of Income and Program Participation (SIPP), including a description of its content, discussion of data collection and processing procedures, and explanation of how to use the SIPP data files made available for public use.

This Guide is still being developed. In future updates to the Guide, more will be written about the methodology of the survey: specifics of sample selection, field procedures, and processing steps including editing and imputation. Further, longitudinal products currently under development will be described. The survey itself has been designed as a flexible instrument, and its contents may be adapted in the future to meet changing needs and conditions or improvements may be made to the data files. Therefore, updates to this Guide will be issued periodically to provide current information and progressively broaden the scope.

To make sure you have the latest updates to the Users' Guide, you can write to Customer Services, Data User Services Division, Bureau of the Census, Washington, DC 20233, or call 301-763-4100.

CONTENT OF THE USERS' GUIDE

The Users' Guide includes the following seven chapters which discuss general concepts and currently available cross-sectional products and two appendices which provide additional information.

Concepts

- Chapter 1, the Introduction, gives an overview of SIPP, beginning with an explanation of the survey's purpose, and its origin in the Income Survey Development Program. This chapter also lays the foundation for other chapters by citing key concepts in content, design, and products associated with SIPP.

- Chapter 2 explains Survey Design, laying out the pattern of data collection. This is important for an understanding of the context of content variables, as discussed in the next chapter, and the structure of SIPP data files, as discussed in chapter 4.

- Chapter 3, Survey Content, describes the types of variables collected in the survey. The data collection instruments are discussed as well. All income sources, programs, and assets covered in SIPP are listed. The chapter also discusses the more specialized "topical modules" associated with specific collection periods.

Cross-Sectional Products

- Chapter 4 discusses the Structure of Cross-Sectional Public-Use Microdata Files from SIPP, explaining the two types of file structures that the Census Bureau makes available. This chapter, used in concert with the technical documentation for the file in use, will enable the user to begin processing the data.

- Chapter 5, Use of Cross-Sectional Files for Estimation and Analysis, guides the user in obtaining cross-sectional estimates. The chapter explains sample weights and the development of estimates for various types of analytic units. The use of a partial SIPP sample is discussed since a full sample of interviews may not be available in the same file together for a given calendar period. Chapter 5 also discusses how the longitudinal nature of the survey, with people entering and leaving the sample and moving from one address to another, affects the derivation of cross-sectional estimates. The chapter further points out the uses and limitations of the geographic information shown in the files.

- Chapter 6, Linking Waves, describes how successive cross-sectional files can be linked to form longitudinal files for analysis. Aside from the basic procedure, the chapter guides the user through the conceptual issues involved in both construction of linked cross-sectional files and their use.

- Chapter 7, Assessing Reliability of SIPP Data, contains the statistical background and formulas to measure the sampling error of the estimates the user derives. Use of an approximation method, calculated with standard error tables, is discussed, as well as a method using half-sample codes on the microdata files. Nonsampling error is also addressed, with particular reference to imputation for missing responses and editing of inconsistencies in the data.

Appendix A is an annotated bibliography of available papers on SIPP.

Appendix B summarizes ordering information for each of the files and reports currently available.

RELATIONSHIP OF THIS GUIDE TO OTHER REFERENCES

This Guide has been designed to complement the technical documentation to the various microdata files, which gives more specific information on content and format of the files. The user may also wish to obtain the SIPP Interviewers Manual for further insight into field procedures and interpretation of certain data items. Furthermore, the user should obtain complete questionnaires for each file used because the content varies over time. A facsimile of the questionnaire is included in the technical documentation which is available from the Data User Services Division, Bureau of the Census, Washington, DC 20233.

Table of Contents

Preface

List of figures

Chapter 1 - INTRODUCTION

- Origins of SIPP
- Income Survey Development Program
- Purposes of SIPP
- Design of the Survey
- Products
- Content of SIPP
- Use of the Data
- Reliability of SIPP Data
- References

Chapter 2 - SURVEY DESIGN

- Key Concepts
- Organizing Principles of SIPP
- Data Collection Units
- Field Procedures
- References

Chapter 3 - SURVEY CONTENT

- Data Collection Instruments
- Information on the Control Card
- Core Questionnaire
- Topical Modules
- References

Chapter 4 - STRUCTURE OF CROSS-SECTIONAL PUBLIC-USE MICRODATA FILES

- Alternate File Structures
- Structure of the SIPP Complex File
- Structure of the Rectangular File
- Reading the Data Dictionary
- Technical Conventions of the SIPP Tapes
- Other Aspects of Working With SIPP Files
- References

Chapter 5 - USE OF CROSS-SECTIONAL FILES FOR ESTIMATION AND ANALYSIS

- Identification System
- Person As the Unit of Analysis
- Households and Families As Units of Analysis
- Preparing Estimates for Different Periods of Time
- Preparing Estimates for Different Geographic Units
- References

Chapter 6 - LINKING WAVES

Conceptual Issues
Procedure for Linking Successive Waves
Linking Nonsuccessive Waves
References

Chapter 7 - ASSESSING RELIABILITY OF SIPP DATA

Sampling Error
Nonsampling Error
References

Appendices

A. Additional References
B. What's Available

FIGURES

Figure

2.1 Overlapping Panels
2.2 Interviewing Schedule for Waves 1-3 of the 1984 Panel
2.3 1984 SIPP Panel Reference Periods
2.4 Illustration of the "Following Rules"

3.1 Types of Income Recorded in SIPP
3.2 Topical Modules for the 1984 Panel
3.3 Topical Modules for the 1985 Panel
3.4 Topical Modules for the 1986 Panel
3.5 Topical Modules for the 1987 Panel

4.1 Hierarchical File Structure
4.2 Month-to-Month Changes Shown in a Complex File
4.3 Components of Rectangular Record
4.4 Summary of Records on the SIPP Complex File
4.5 Sequence of Records on the Complex File: 1984 Panel, Wave 2 Example
4.6 Record Sequence on the Complex File: 1985 Panel, Wave 4 Example
4.7 Geographic Codes Used in SIPP
4.8 Rectangular File Records for the Hypothetical Example
4.9 Excerpt From Data Dictionary of Wave 1 Rectangular File

5.1 Identification System for the Sample Unit Depicted in Figure 2.4
5.2 Source of Questionnaire Information: Reported VS. Imputed
5.3 Statistical Properties of Person Weights on 5 Waves of the 1984 Panel
5.4 Example of Weighted Data
5.5 Composition for the Hypothetical Sample Unit in Wave 8
5.6 Multiple Food Stamp Unit
5.7 1984 SIPP Panel Reference Periods
5.8 Factors for Monthly Data: Wave 1, 1984 Panel
5.9 Factors to Be Applied to Monthly Data to Obtain Monthly Estimates
5.10 Metropolitan Subsample Factors: 1984 Panel

6.1 Possible Transitions Between Two Consecutive Waves, N and N+1
6.2 Unweighted Counts of Type Z Nonrespondents Across Waves 1, 2, 3 of the 1984 Panel, Rotation Groups 1, 2, and 3
6.3 Variables Needed to Identify Individuals and to Determine the Classification of Each Record
6.4 Cases Which May Be Excluded From Longitudinal Analyses
6.5 Verification Statistics for Matching Successive Waves in the 1984 Panel

Figure

7.1 SIPP 1984 Generalized Variance Parameters for Estimated Numbers and Percentages of Households and Persons
7.2 Correlations for Monthly and Quarterly Comparisons
7.3 Distribution of Monthly Household Income Among Persons 25 to 34 Years Old
7.4 Sample Sizes in Addresses by Wave and Interview Status, 1984 Panel
7.5 Illustration of an Imputation Flag on Wave 3, 1984 Panel

CHAPTER 1 - INTRODUCTION

ORIGINS OF SIPP

American society is marked by complexity and change. Our post-industrial economy, our prevailing philosophy of individual fulfillment, and our belief in the freedom to pursue individual goals within the bounds of social responsibility have produced a highly mobile society.

Adding to this complexity are a whole host of government programs designed to transfer portions of the national product to individuals whose well-being would otherwise fall below some loosely defined minimum level. Originating in the 1930's, some programs proved to be temporary, such as the employment of the jobless by the Works Progress Administration, now defunct; others, such as Social Security, unemployment insurance, and Aid to Families with Dependent Children, are still in existence. Workmen's compensation laws were passed even earlier. Federal aid to the disabled, authorized in 1950, continues as a basic support program.

Another set of programs, mostly directed at eliminating poverty and its surrounding conditions, originated during the 1960's and 1970's. These included expansion of public assistance, food stamps, subsidized school lunches, Medicaid, Medicare, and housing subsidies.

Private transfer programs also have a long history. Aside from company pension plans, families share resources to help support their elderly and child members as well as others unable to work. Private charities also provide cash and noncash assistance, though on a much smaller scale than the Federal Government.

The effectiveness of transfer programs, particularly those administered by the Federal Government, has long been a subject of public debate. (See Batchelder (1971) for a review of government transfer programs and the arguments surrounding them in the 1970's.) Attempts to evaluate these programs have been frustrated by a lack of critical data on cash and noncash income, asset components of wealth, subannual program participation patterns, and the dynamics of household relationships.

Until the advent of the Survey of Income and Program Participation (SIPP), the major source of data on income and program participation has been the March Income Supplement of the Current Population Survey (CPS). The CPS, however, is primarily designed to obtain information on employment and unemployment. Because income measurement was never the primary purpose of CPS, it leaves certain gaps in this area. For example, in CPS respondents are asked in March to recall their income during the preceding calendar year. Many have difficulty in remembering sources such as property income or irregular income over the year-long reference period. Also because of the annual reporting period, CPS does not capture the impact of household moves and changes in composition during the year, nor does the survey explicitly measure part-year periods of program participation. Further, in CPS there is no room to collect data on assets and liabilities which are needed to measure more completely a household's economic status and eligibility for program benefits. To add these items to the CPS questionnaire would dilute the main purpose of the survey and unduly increase respondent burden.

Gaps also exist in CPS data on program participation. Without careful probing, respondents can easily confuse one income security program with another, resulting in faulty or incomplete data. CPS does not collect information sufficient to fully examine eligibility for and participation in various Federal, state, and local assistance programs. Eligibility for many of these programs is based on monthly rather than annual income. Federal program managers and researchers from many areas needed data for a shorter accounting period.

INCOME SURVEY DEVELOPMENT PROGRAM

Awareness of these limitations led to support for a new survey, and the Income Survey Development Program (ISDP) came into being as a joint program of research and development by the Department of Health and Human Services and the Census Bureau (see Hunt, 1985).

Extensive research was undertaken over a 5-year period (1975 to 1980) to improve the measurement and collection of income, program, and wealth data. Much of the work centered around four experimental field tests, conducted to examine different concepts, procedures, questionnaires, and recall periods. Two tests, the 1977 Site Research Test and the 1978 Research Panel, while relatively small, demonstrated that many new ideas and methods were feasible. This work laid the basis for the largest and most complex test, the 1979 Research Panel.

For many purposes the 1979 Research Panel was a dry run of SIPP. It was a longitudinal survey of a nationally representative sample of adults in 7,500 households. Persons in the households were contacted in early 1979 and asked questions about their income, labor force participation, and other characteristics. These persons were recontacted every 3 months to supply information on themselves and others with whom they resided, covering the months since the previous interview. The basic information was collected over six interviews. The second and subsequent interviews covered additional topics such as medical care, work history, marital history, child care, and pension coverage.

Although the survey was used primarily for methodological purposes, the sample size was large enough to provide reliable national estimates of many characteristics of interest. Examples of ISDP use include analyses of food stamp participant turnover, wealth and income of the elderly, and the effects of work experience on earnings.

Information and discussion of uses of ISDP can be found in references listed in appendix A, and in Long, et. al. (1986), David (1983), and Yeas and Lininger (1981). Public-use microdata files, questionnaires, and documentation of ISDP are available through the National Technical Information Service, 5385 Port Royal Road, Springfield, VA 22161.

PURPOSES OF SIPP

The Survey of Income and Program Participation (SIPP) is expected to become a major source of information on the demographic and economic situation of the United States. Analysis of the data will provide a better understanding of the distribution of income, wealth, and poverty in the society, and of the effects of Federal and state programs on the well-being of families and individuals.

SIPP's purposes, like those of ISDP, are to improve the measurement of the economic situation of persons, families, and households in the United States, and to provide a tool for managing and evaluating government transfer and service programs.

These are the major goals of SIPP:

- To improve accuracy in reporting and classifying income sources.
- To obtain subannual information on income recipiency and program participation.
- To examine interactions among transfer programs, labor force participation, and living arrangements.
- To obtain sufficient information to improve the simulation of eligibility under the major means-tested cash and in-kind transfer programs.
- To obtain improved measures of assets and liabilities.

To improve estimation of income, information is collected on many types of earned, unearned, and asset income sources. Another method used to improve estimation of annual income is the measurement of monthly variations in such contributing factors as household structure, the determinants of program eligibility, and actual participation. Greater understanding of the dynamics of social change and the effectiveness of public programs should result.

SIPP's longitudinal features enable the analysis of selected dynamic characteristics of the population such as changes in poverty status and associated events. A better understanding of the interrelationships between poverty and social and employment characteristics will result, allowing more efficient design and better targeting of transfer programs.

In addition to income sources, SIPP collects information on asset ownership. Analysts can gain, thereby, a better picture of the total economic resources available to families and persons. Such an understanding has implications for transfer program design and for the creation of better social indicators. One of the most important reasons for conducting SIPP is to gather more precise information on participation in transfer programs. Fragmented data on individual programs were available heretofore, but, with SIPP, analysts can examine multiple recipiency of means-tested transfers (such as Aid to Families with Dependent Children--AFDC) and nonmeans-tested transfers (such as Social Security). For an example of multiple transfer program participation, see Weinberg (1985). Data on private transfers (such as company pensions or private health insurance) are also available to round out the recipiency picture.

One family of uses involves research on the efficacy and effects of transfer programs. The data could help answer questions such as these:

- How would changes in eligibility rules or benefit levels directly affect recipients and program costs?

- How would changes in the eligibility rules affect the program target population, i.e., those eligible to receive benefits?

- How does income gained by other household members affect labor force participation and reasons for not working?

- How do wealth and income patterns differ for various age, sex, and race groups?

Since SIPP is a longitudinal survey capturing changes in household and family composition over a 2 1/2-year period, analyses such as the following are possible:

- What factors affect change in household and family structure and living arrangements?

- What are the interactions between changes in the structure of households and families and the distribution of income?

- What effects do changes in household composition have on economic status and program eligibility?

- What are the primary determinants of turnover in programs such as food stamps and AFDC?

SIPP, then, can be used in cross-sectional or longitudinal analysis, to pursue social research as well as that directed toward improving public policy.

DESIGN OF THE SURVEY

SIPP is a multi-panel longitudinal survey of adults, measuring their economic and demographic characteristics over a period of 2 1/2 years. The adults followed in each panel of the survey are determined by a nationally representative survey of households in the civilian noninstitutionalized population. The first panel began in October 1983 with the adults in 19,878 interviewed households. The second and subsequent panels begin in February of each calendar year, resulting in concurrent administration of the survey in multiple panels. Persons selected into the SIPP sample continue to be interviewed once every 4 months over the 2 1/2-year life of the panel. If persons initially interviewed move from their original address to another address, they are interviewed at the new address. (There are certain exceptions, discussed in chapter 2.) If, after the first interview, the respondents move into households with other people not previously in the survey, the new people are interviewed as long as they continue living with the first interview respondents. Over a series of interviews a record is built up for most of the sample members of their moves to different locations and of how their households change. Thus, SIPP provides cross-sectional views of respondents' lives at discrete points in time, as well as a longitudinal history of changes in their economic circumstances and household relationships.

To facilitate field procedures, each sample panel is divided into four random subsamples, each representative of the Nation, called "rotation groups"; each rotation group is interviewed in a separate month. Four rotation groups thus

comprise one cycle or wave of interviewing for the entire panel. (There is one exception discussed in chapter 2.) At each interview, respondents are asked to provide information covering the 4 months since the previous interview. This 4-month span is the "reference period" for the interview.

Chapter 2 contains a more detailed description of SIPP design.

PRODUCTS

A number of publications and computerized data files are generated from the information collected in SIPP. Currently, the available products are cross-sectional; that is, featuring data for only a single collection period. In the future, longitudinal products which will permit examination of changes or trends over several interviews will be available.

The Census Bureau publishes reports entitled "Household Economic Studies" available in Current Population Reports, Series P-70. The first six reports in the P-70 series were quarterly reports entitled "Economic Characteristics of Households in the United States." These reports provided aggregate values for a variety of labor force, income, and household composition measures in terms of average monthly estimates as observed during the calendar quarter. With the release of P-70, No. 7, "Household Wealth and Asset Ownership: 1984" (issued in July 1986), these reports are no longer quarterly and are released throughout the year under different subject-matter titles. They will contain data on income and earnings, including income derived from public programs, and data on a wide variety of other related topics such as assets, liabilities, health insurance coverage, child care, taxes, etc. Plans for additional printed products are in development at this writing.

A series of working papers on SIPP is also available. Appendix B describes the papers and contains ordering information for them as well as the reports mentioned above.

Compared to the vast analytical scope of SIPP, the contents of the printed reports are necessarily limited. Fortunately, researchers also have access to public-use microdata from the survey. These consist of unaggregated records for individual respondents, edited to remove information which might be used to determine the identity of any respondent. These microdata allow a user to construct tables not included in the published reports, to employ the data to determine the population eligible for various tax and transfer programs, or to perform a wide variety of other statistical operations.

Cross-sectional microdata files from SIPP are issued on a wave-by-wave basis. Files from waves which do not collect topical modules, such as the first two interviews in the 1984 Panel, consist only of core data. Files from other waves also include topical questions.

Chapter 4 discusses the structure of the cross-sectional files. The microdata are available in two structural formats, rectangular and relational (complex). Cross-sectional files are released with each wave of data. These can be linked to create longitudinal files. Note, however, that the cross-sectional files for each wave have been produced independently from the other waves so that linked cross-sectional files may not be straightforward to use for longitudinal analysis. This topic is discussed in chapter 6 of the Users' Guide.

Public-use microdata files from SIPP available at this writing are summarized in appendix B.

CONTENT OF SIPP

Information collected in SIPP falls into two categories. The "core" of the survey includes questions asked at every interview. Other questions produce in-depth information on specific subjects and are asked at only one or two interviews. These questions are found in "topical modules" that are added to the questionnaire for many of the interviews and in the baseline interview conducted first.

The core questions cover demographic characteristics; labor force participation; program participation; amounts and types of earned and unearned income received, including transfer payments and noncash benefits from various programs; and asset ownership. Questions on other subjects such as private health insurance and pension receipt are also included in the core. Furthermore, selected items not considered to be core data in the survey are integrated with these core questions in the first round of interviewing. These include items such as reasons for receiving social security payments which are only asked once per individual.

Topical modules probe in greater detail about particular social and economic characteristics and personal histories. These include such topics as assets and liabilities held, school enrollment, marital history, fertility, migration, disability, and work history. These topical modules typically contain information on events in the past or characteristics that tend to change slowly. Thus, these questions do not need to be asked at every interview.

Most core data are measured on a monthly basis, making detailed time series analysis possible. Some core information is measured every 4 months at the time of the interview. Data from a topical module or data not considered to be core data at the time of the first interview are available only once or twice during the panel, at best yielding estimates of annual change.

Chapter 3 describes the content of SIPP in detail.

USE OF THE DATA

Although the public-use cross-sectional data products may first appear fairly straightforward to analyze, there are a number of issues to be considered in the analysis. First, even within each cross-sectional file, the population represented is not static. People enter and leave the sample and change their relationships from month to month. Second, the survey employs a complex sample design and multiple weights have been assigned to each individual for use in the analysis. Third, the publicly available data have been altered to some extent to preserve the confidentiality of the respondent. Finally, the derivation of statistics for calendar time periods is not straightforward.

These topics are discussed in chapter 5.

LINKING WAVES

As noted previously, SIPP is designed to permit users to analyze the dynamics of the population such as the factors affecting family formation and dissolution and the determinants of program turnover. However, to do so requires longitudinal microdata products which are not developed immediately upon collection of the data by the Census Bureau. The user can elect to perform longitudinal analysis in advance of the longitudinal products, though, by linking successive cross-sectional files. The user may also wish to link topical information in one round of interviewing with core or topical data collected in another round of interviewing. In order to assist users in this endeavor, chapter 6 is devoted to the conceptual and practical issues to be faced in linking the public-use microdata files.

RELIABILITY OF SIPP DATA

Estimates from SIPP, like other surveys, are subject to errors of two different kinds. The first type of error, called sampling error, results from the fact that outcomes from the SIPP sample may differ from those which would be obtained had the entire population been surveyed. The second type of error, called nonsampling error, results from undercoverage of the population, nonresponse on the part of the sample members, and errors made during data collection and processing. Chapter 7 provides an overview of how to calculate sampling error in SIPP, and further discusses the extent and sources of nonsampling error.

REFERENCES

Batchelder, P. A. The Economics of Poverty. New York: John Wiley & Sons, Inc., 1971.

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

Hunt, M. Profiles of Social Research. New York: Russell Sage Foundation, 1985 (pp. 98-154).

Long, S., K. Beebout, and F. Skidmore. Food Stamp Research: Results From the Income Survey Development Program and the Promise of the Survey of Income and Program Participation. Washington, D.C.: U.S. Department of Agriculture, Food and Nutrition Service, 1986.

Weinberg, D. H. "Filling the 'Poverty Gap': Multiple Transfer Program Participation," Journal of Human Resources, 1985 (pp. 64-89).

Yeas, M. and C. Lininger. "The Income Survey Development Program: Design Features and Initial Findings," Social Security Bulletin, Vol. 44, No. 11 (1981).

CHAPTER 2 - SURVEY DESIGN

This chapter describes the design features of the Survey of Income and Program Participation (SIPP). Understanding the survey design provides a basis for understanding the data and working with the data files. The chapter first introduces some of the main design concepts, then explains in greater detail how the survey is organized.

KEY CONCEPTS

The SIPP is a longitudinal survey of adults in households obtained from a multi-stage stratified sample of the noninstitutional resident population of the United States. It is a multi-panel survey with a new sample (panel) introduced at the beginning of each calendar year. The initial selection of households into the survey is done according to a sample selection methodology similar to that used for the Current Population Survey.[1] The definition of which adults actually form the sample is provided below.

The first panel of SIPP (the 1984 Panel) was initiated in October 1983. Of the approximately 26,000 addresses designated for the survey, about 5,000 were found to be ineligible. Another 1,000 households could not be interviewed because they refused, could not be found at home, were temporarily absent, or were otherwise unavailable. These households were dropped from the survey. Thus, the actual sample for the 1984 Panel consisted of the adults in 19,878 interviewed households.

Initial interviews were conducted in each of the 19,878 households selected into the panel during the period October 1983 through January 1984. An attempt was made to reinterview adults in these households at 4-month intervals thereafter.

In each interview information is collected for original sample adults and other individuals with whom they reside. The data obtained include labor force status, program participation, and income during the 4 calendar months preceding the interview, i.e., the "reference period." To spread out the work load for interviewers, the sample is divided into four "rotation groups," and a different rotation group is interviewed every month. Each round of interviews is referred to as a wave.

Adults in the 1984 Panel sample continue to be interviewed for up to nine waves, thus collecting data for up to 3 years (36 months). For subsequent panels, interviews start in February of the panel year, and most of the sample members will remain in the survey for eight interviews, collecting data covering 2 years, 8 months. Thus, after 1984, the panels overlap for part of their duration.

The primary focus of SIPP is adults, i.e., persons 15 years old or older in the initial household sample. These individuals, called "original sample persons," are followed over the life of their panel, even if they move from the original housing unit (with certain exceptions discussed in "Following Movers" in the "Field Procedures" section). SIPP also collects data on all other persons,

---

[1] More detail on sample selection is given in chapter 5.

including children, who live with or move into households with any of the original sample persons. These individuals, however, are not followed when they no longer reside with original sample persons.

The rest of the chapter explains the design more fully and illustrates the concepts with examples.

ORGANIZING PRINCIPLES OF SIPP

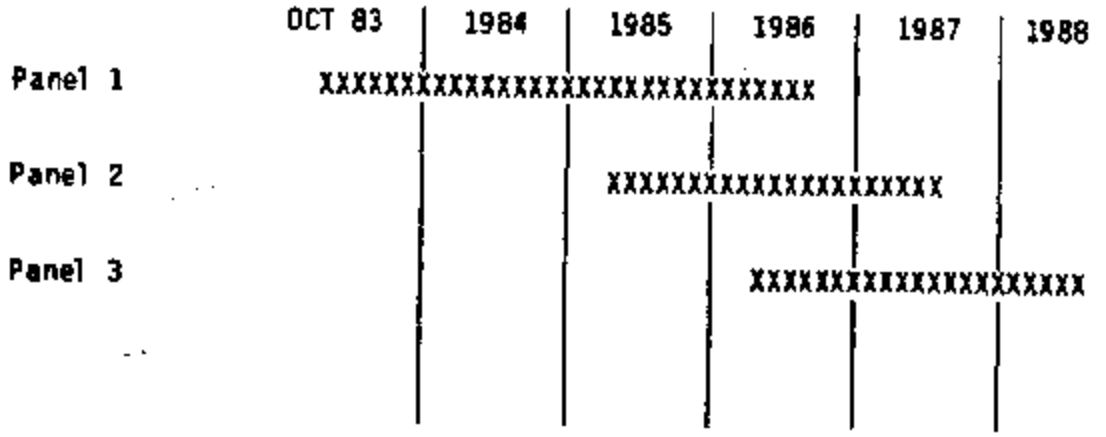
Panels

Figure 2.1 shows the overlapping panel design for the 1984, 1985, and 1986 panels. The 1984 Panel remained in the survey until July 1986. The 1985 Panel began in February 1985 and will remain in the survey until August 1987. The 1986 Panel began February 1986 and is scheduled to end August 1988.

Each panel, then, remains in the survey for about 2 1/2 years. For most respondents, the last interview comes 2 years and 4 months after the first interview. Since the first interview asks for information covering the preceding 4 months, the total period covered by the survey is 2 years, 8 months. The 1984 Panel has an extra 4 months of interviewing because it began in October 1983.

While the 1984 Panel at the start consisted of 19,878 interviewed households, this was reduced by about 3,000 after the start of 1985.[2] Subsequent panels will be smaller right from the beginning, including adults in 13,349 households for the 1985 Panel and 11,513 for the 1986 Panel. During 1985, when the 1984 and 1985 Panels overlapped, the SIPP sample consisted of about 30,000 interviewed households. From February to July 1986, when three panels were in the survey, data were collected for about 37,000 households.

FIGURE 2.1 - OVERLAPPING PANELS



[2]The differential subsampling which reduced the size of the 1984 Panel is discussed briefly in chapter 5. With the combination of the reduction in sample size for budgeting purposes, sample attrition, and changing household composition, the number of interviewed households in Wave 5 of the 1984 Panel was 16,978.

The overlapping design is important for both cross-sectional and longitudinal analysis. Data from multiple panels result in a larger number of cases, reducing sampling error. Greater reliability is thereby achieved in estimates of change and cross-sectional estimates. Continuation of each panel over a period of 2 years and 8 months makes estimates of year-to-year change more reliable since they can be based in part on a direct comparison of the same households across 2 years. This panel duration also allows collection of income and tax data for 2 calendar years. An annual topical module on income sources, amounts, and taxes is scheduled for the wave of interviews occurring in late spring and early summer in the second and third year of each panel--timed to take advantage of the record-gathering many individuals do in compiling their income tax returns.

The design has other advantages that pertain to the types of data collected and the uses of the information. Panels are geared to the calendar year because some kinds of information are obtainable only on a calendar-year basis, and most past survey work has gathered primarily calendar-year data. At the same time, monthly estimates derivable from SIPP are useful in analyzing participation in many programs that rely on monthly income and household composition in determining eligibility.

Interviewing Schedule

The entire sample is interviewed by personal visits to original sample persons once every 4 months. In general, each group of interviews with the full sample is termed a wave. Wave 1 of the 1984 Panel, for example, began in October 1983. Four months later, Wave 2 began in February 1984.

The 1984 Panel includes nine waves of interviews. Subsequent panels include eight waves.

SIPP data collection is a continuous rather than a periodic process, so that the field staff remains dedicated to SIPP and gaps in the workload are minimized. This continuity is achieved by dividing the SIPP sample into four random subsamples of approximately equal size, called rotation groups. One rotation group is interviewed each month in succession, resulting in coverage of the entire sample in 4 months; e.g., for Wave 1, 1984 Panel, interviews were conducted from October 1983 to January 1984. Rotation Group 1 households were interviewed in October; Rotation Group 2, in November; Rotation Group 3, in December; and Rotation Group 4, in January.

In theory, the combination of continuous interviewing and 4-month interview cycles produces a balanced pattern of interviews begun every 4 calendar months with a new questionnaire. However, due to budget restrictions, this balanced pattern does not actually occur. In each panel (through the 1986 Panel), there is at least one wave of interviewing that consists of fewer than four rotation groups. This does not mean that original sample adults in one rotation group have a gap in the data. Instead, the interviewers arrive at sample members' address at the regular 4-month interval with a different questionnaire. To be more explicit, the interviewing schedule used in the early waves of the 1984 Panel is depicted in figure 2.2.

Note that even though Rotation Group 4 was "dropped" from Wave 2, there was no gap in the data collected. There are several implications of this procedure.

First, selected cross-sectional files in each panel will contain fewer observations than the other cross-sectional files. For example, Wave 2 of the 1984 Panel only has three-fourths of the sample. Second, selected topical information will not be collected for the full sample. For example, Wave 8 in the 1984 Panel has a reduced sample (Rotation Group 3 received the Wave 9 questionnaire instead of the Wave 8 questionnaire--see chapter 5 for more information) and it included topical modules covering support for nonhousehold members; work-related expenses; marital, migration, and fertility history; and household relationships. Finally, the length of the period over which original sample adults can be followed is reduced for selected rotation groups. For example, Rotation Groups 3 and 4 will have 4 fewer months of data in the 1984 Panel than Rotation Groups 1 and 2. The 1985 and 1986 Panels also contain one wave shorter than the others. To avoid the confusing nomenclature by which Rotation Group 4 seems to be interviewed out of sequence during most waves, the labelling

FIGURE 2.2 - INTERVIEWING SCHEDULE FOR WAVES 1-3 OF THE 1984 PANEL

Wave	Interview Date	Rotation Group Interviewed	Questionnaire Administered	Reference Period
1	October 1983	1	1	June 1983-Sept. 1983
	November 1983	2	1	July 1983-Oct. 1983
	December 1983	3	1	Aug. 1983-Nov. 1983
	January 1984	4	1	Sept. 1983-Dec. 1983
2	February 1984	1	2	Oct. 1983-Jan. 1984
	March 1984	2	2	Nov. 1983-Feb. 1984
	April 1984	3	2	Dec. 1983-Mar. 1984
3	May 1984	4	3	Jan. 1984-Apr. 1984
	June 1984	1	3	Feb. 1984-May 1984
	July 1984	2	3	Mar. 1984-June 1984
	August 1984	3	3	Apr. 1984-July 1984

of rotation groups in the 1985 Panel is slightly different. When Wave 1 began in February 1985--instead of numbering the Rotation Groups 1, 2, 3, and 4--they were labelled 2, 3, 4, and 1. Wave 2 began in June 1985 with Rotation Group 2, followed by Rotation Group 3, then 4. (Wave 2 lasts only 3 months.) Wave 3 and subsequent waves then proceed in the natural order of Rotation Groups 1, 2, 3, and 4. In the 1986 and subsequent panels, Wave 3 is the short wave (lasting only 3 months) rather than Wave 2. Rotation group labelling in these later panels is similar to the 1985 Panel.

Reference Periods

The final element of the SIPP survey design is the "reference period" for each interview. Data are collected for each of the 4 calendar months preceding the interview which are called "reference months." Thus, except for Wave 1, the information covers the calendar month during which the previous interview was conducted and the 3 intervening months. The short SIPP reference period is designed to reduce respondents' difficulties in recalling specific information, resulting in greater accuracy.

The relationship of reference periods and reference months to rotation groups and waves is illustrated in figure 2.3, which shows the structure of the 1984 Panel. Wave 1 interviews occurred in October, November, and December 1983, and January 1984. The reference period for the October interviews, conducted with respondents in Rotation Group 1, provided information covering June through September, as well as information pertinent at the time of interview. To illustrate the availability of time-specific data in SIPP, the first items in the SIPP interview (after gathering basic demographic characteristics) ask whether the respondent had a job or business at any time during the previous 4 months, and, if he or she experienced unemployment, in which specific weeks he or she spent time looking for work or on layoff. Later in the interview, persons with jobs are asked to report earnings in each of the 4 reference months. Individuals who report owning assets are asked to report total income from each asset for the 4-month period (this lump sum amount is later allocated to months). Some information on the data file, then, is specific to the entire reference period; other items are month-specific; and some items are week-specific.

During Wave 2 of the 1984 Panel, the respondents in Rotation Group 1 were interviewed in February 1984. Information collected in that interview covered the 4 months of October 1983 to January 1984. Wave 1 for this rotation group had covered June to September 1983. Thus, a continuous record of data is built up over successive waves of interviewing, covering every month during the life of the panel. Note that some information pertaining to October, the interview month of Wave 1, was collected twice--once in Wave 1 and once in Wave 2.

In figure 2.3, the columns portray reference periods and months. Years and quarters are explicitly labelled, while individual months are identified only by their first letter. It can be seen that in the second quarter of 1983, data are present for June only for Rotation Group 1, thereby including only one-fourth of the sampled households for the 1984 Panel. It can be seen that two rotation groups were interviewed with regard to July and three with regard to August 1983. September 1983 is the first month for which data were collected from all four rotation groups. October 1983 illustrates the fact that, to obtain data for a given calendar month, it is frequently necessary to work with two

FIGURE 2.3 - 1984 SIPP PANEL REFERENCE PERIODS

Wave	Rotation Group	Interview Month	Reference Period: 1983 2nd Qtr	1983 3rd Qtr	1983 4th Qtr	1984 1st Qtr	1984 2nd Qtr	1984 3rd Qtr	1984 4th Qtr	1985 1st Qtr	1985 2nd Qtr	1985 3rd Qtr	1985 4th Qtr	1986 1st Qtr	1986 2nd Qtr
1	1	Oct	J	JAS											
	2	Nov		JAS	O										
	3	Dec		AS	ON										
	4	Jan		S	OND										
2	1	Feb			OND	J									
	2	Mar			ND	JF									
	3	Apr			D	JFM									
3	4	May				JFM	A								
	1	Jun				FM	AM								
	2	Jul				M	AMJ								
	3	Aug					AMJ	J							
4	4	Sep					MJ	JA							
	1	Oct					J	JAS							
	2	Nov						JAS	O						
	3	Dec						AS	ON						
5	4	Jan						S	OND						
	1	Feb							OND	J					
	2	Mar							ND	JF					
	3	Apr							D	JFM					
6	4	May								JFM	A				
	1	Jun								FM	AM				
	2	Jul								M	AMJ				
	3	Aug									AMJ	J			
7	4	Sep									MJ	JA			
	1	Oct									J	JAS			
	2	Nov										JAS	O		
	3	Dec										AS	ON		
8	4	Jan										S	OND		
	1	Feb											OND	J	
	2	Mar											ND	JF	
9	3	Apr											D	JFM	
	4	May												JFM	A
	1	Jun												FM	AM
	2	Jul												M	AMJ

Note that Waves 2 and 8 only contain three rotation groups. For subsequent panels, the figure would be similar except each panel begins in February of each panel year, the rotation groups are numbered differently, only eight waves are conducted, and only one wave contains less than the full sample (Wave 2 in the 1985 Panel and Wave 4 in subsequent panels).

different waves; for October 1983, one needs data from Wave 1 (Rotation Groups 2, 3, 4) and Wave 2 (Rotation Group 1). Full sample data for a particular quarter of a year always require at least two waves and may require three; for example, the third quarter of 1984. For that quarter, July data are needed from Wave 3; July, August, and September data are needed from Wave 4; and September data are needed from Wave 5. (Deriving estimates for particular time periods and procedures for using data from less than the full sample are discussed in chapter 5.)

DATA COLLECTION UNITS

There are four types of data collection units in SIPP--sample units, address units, families, and persons. (Chapter 4 includes a discussion of these units as records in the SIPP data file structure and illustrates these concepts through explicit examples.)

Sample Unit

Each address initially selected into each SIPP panel is called a sample unit. There were 19,878 sample units initially selected into the 1984 Panel. Household composition as of Wave 1 (or, more specifically, the persons 15 years and over found in interviewed Wave 1 households) identified the people to be followed for the duration of the panel. The number of sample units in subsequent panels is smaller, with the actual size fluctuating in response to budget constraints.

Address Unit

This collection unit is a person or group of persons living at the same address at the time of each interview. The address unit may consist of one person living by himself or herself, a group of unrelated individuals, or one or more families.[3]

Sample unit membership is defined as of the Wave 1 interview date in each panel with initially only one address per sample unit. The Wave 1 date is chosen to establish a reference point. If the SIPP interviewer finds in subsequent interviews that one of the original household members 15 years old or over has moved elsewhere, that person's new address is eligible for interview. Thus, it is possible for original sample members to be part of two or more address units over the period of the survey.[4]

In cross-sectional surveys like the CPS, this address unit is referred to as a household. In fact, this is true for Wave 1 of each SIPP panel. However,

---

[3]Persons in noninstitutional group quarters, such as rooming and boarding houses, college dormitories, and convents or monasteries, are eligible for inclusion. They are interviewed if the address is selected into the sample and the group quarters is their usual residence. Thus, college students in dormitories would usually not be interviewed, as their usual residences are with their families.

[4]With a few exceptions, all persons remain within the original sample unit even when they move. The exception occurs when two or more persons from different sample units unite into one household.

it is useful to make a distinction between households and address units when thinking about SIPP in general. Households are defined by month based on persons sharing an address. Households, at the time of the interview, are synonymous with address units; however, households in other months may not be. Monthly household composition is not explicitly collected in SIPP. Instead, it is inferred from the interviews conducted at each address, based on information obtained regarding persons entering and leaving the address.

Users should exercise caution in using SIPP data on addresses. They do not necessarily define a continuous household where the household is viewed as an analytic unit. The issue of how to define households longitudinally as units of analysis has received much attention as illustrated by McMillen and Herriot (1985); Citro, Hernandez, and Herriot (1986); Duncan and Hill (1985); and Citro, Hernandez, and Moorman (1987). However, it has not been resolved absolutely.

On the public-use files, economic and demographic data are summarized for each household in each month. In addition, there is a limited amount of demographic information summarized for the address unit or the household at the time of interview. Some information such as tenure is collected only once per address. This is allocated to monthly households based on the address itself.

Family

The family is defined by the Census Bureau as two or more persons who are living together and related by blood, marriage, or adoption. On the public-use files, economic and demographic data are summarized for each primary and unrelated subfamily unit and each unrelated individual. A primary family is the family containing the household reference person; an unrelated subfamily is a family which does not contain the reference person nor anyone related to the reference person.

SIPP also notes the presence of related subfamilies--or families within the primary family. A daughter and husband living with the daughter's parents would constitute a subfamily, and the same variables would be summarized for the subfamily as for the entire primary family. Note that, although data for the daughter and husband are summarized separately, they are also counted in the summary information for the primary family.

Family composition may change from month to month. Families may dissolve (as through divorce) or be created (as when two individuals marry) from one month to the next. SIPP is a unique data source in that it allows the construction of family relationships on a monthly basis, along with other monthly information such as employment and income. In Wave 1, the monthly composition variables all reflect the composition as of the interview date, so the monthly composition is the same for each reference month in Wave 1. However, in all other waves composition varies by month.

Person

SIPP collects data on all persons 15 years old and over in the original sample household, as well as additional persons who live with them in that or any future household. An additional person remains in the survey until he or she no longer shares a household with the original sample person. Thus, an attempt

is made to collect complete data on the living arrangements of adults originally selected into the sample for the full length of the panel. More information on following these original sample members when they move is found in the next section.[5] With the exception of some children, no attempt is made to follow other individuals when they no longer reside with original sample adults. The exceptions are children under 15 in Wave 1 households who remain with original sample adults until after attaining age 15 and then move. (They must have had their 15th birthday by the last day of the reference period.) These persons are treated as original sample adults even though they were not adults as of the Wave 1 interview.

FIELD PROCEDURES

Interviewing

Field procedures for Wave 1, the initial interviews with SIPP households, are similar to those for other Census Bureau surveys. Interviewers list specific addresses of living quarters either prior to or at the time of the interview. They verify these addresses during the interview and determine if additional housing units (such as a basement apartment) are located at the address. In an interviewed household, the interviewer compiles a household roster, listing all persons living or staying at the address. The interviewer then decides whether each person is a household member by using a set of rules which were developed to determine if the address is their usual residence as of the date of the interview.

All interviews are by personal visit, though follow-up, if needed, is done by telephone.[6] The interviewer collects demographic information about all persons in the sample, regardless of age, and then completes a SIPP questionnaire for each household member 15 years of age or older. Each person 15 or older is asked to provide information about himself or herself; if someone is not available at the time of the interview, another adult is asked to provide information about that person (providing a "proxy" response). Because of the detailed nature of the questionnaire, respondents may not be able to furnish certain information at the interview, especially items referring to miscellaneous and asset income. Interviewers arrange to telephone for answers to these "callback" items if the respondents are willing. Callbacks are also made to correct inconsistencies found during questionnaire editing. Telephone interviewing is permitted only to obtain missing information, to interview persons who will not or cannot participate otherwise, or to interview persons who have moved to an area more than 100 miles from any area normally covered by SIPP interviewers (i.e., selected Primary Sampling Units--PSU's).

---

[5] An identification number is associated with each person in the survey. This number allows the user to associate each person with his or her household and to follow movers. The system is described in chapter 5.

[6] From August to November 1986 and February to April 1987, an experiment was conducted to test the feasibility of telephone interviewing. Half of the 1986 Panel was interviewed by phone in three waves of the panel. In Wave 2, it was done for two rotation groups (4 and 1); in Wave 3, two rotation groups (2 and 3); and in Wave 4, three rotation groups (2, 3, and 4).

At the next interview, the interviewer updates the household roster for each sample unit. Additional persons are listed who may have joined the household, such as a new spouse or baby, and the dates they entered. Persons who have left the household and the dates they left are also listed. For those remaining at the same address, if any, the interviewer verifies that certain previously collected information still applies and completes the new questionnaire for each person 15 years or older. Full information is also collected from new household members.

Following Movers

If some but not all sample household members have moved since the last interview, the interviewer completes interviews at the original address and also obtains the new addresses of the individuals who moved. Movers are interviewed at their new address(es), along with anyone else they are living with at that time.

If no original sample members live at an address where a previous interview was conducted, no information is collected from the new occupants of the sampled address. (Out of a total of 16,876, there were 833 such addresses in Wave 2 of the 1984 Panel.) Since SIPP is person-based, the interviewer must attempt to find out the new address(es) of the original sample members. They are then interviewed at their new address(es). (As a result of both new household formation and moves of entire sample households to new addresses, 1,226 addresses were added in Wave 2 of the 1984 Panel.)

Several sources may be used to identify new addresses of movers. At the first interview, the interviewer obtains the name, address, and telephone number of a person who could furnish the new address should the entire household move. Otherwise, interviewers may contact neighbors, employers, mail carriers, real estate companies, rental agents, and postal supervisors. They might even try calling people with the same last name in the local telephone directory. In addition, the Census Bureau mails advance letters to respondents before each interview; if the respondent no longer lives at the address, the post office is requested to provide a forwarding address.[7]

Personal visits are required for all new addresses in or within 100 miles of an area worked by a SIPP interviewer. This expanded area covers 96 percent of the population as it was distributed in 1980. Those moving outside such an area are interviewed by telephone where possible.

An attempt is made to follow original sample adults as long as they remain in the civilian noninstitutionalized population. Initially, for the 1984 Panel, original sample persons who moved outside the United States, entered institutions, or moved to Armed Forces barracks were not followed even if they reentered the domestic noninstitutional population during the life of the panel.

---

[7]For a few waves in the 1984 Panel, change of address forms were left with the respondents with a request that they use these to notify the Census Bureau if they moved. This was discontinued due to a small number of responses.

After returning to the civilian noninstitutionalized population, further information was collected about these persons only if they moved in with other original sample persons still in the survey.[8] If they returned to their original sample unit, they retained their original identifiers; otherwise, they were not linked to their earlier interviews.

Beginning with Wave 6 of the 1984 Panel and Wave 2 of the 1985 Panel, original sample persons institutionalized after the first interview have been tracked but not interviewed during the period of institutionalization. They are interviewed further if they reenter the civilian noninstitutionalized population during the life of the panel. They retain their original identifiers when they reenter and are treated as original sample adults. (Persons moving abroad or into barracks will not be so followed.) This procedure will be followed for the remainder of the 1984 Panel, as well as for the 1985 and subsequent panels.[9] Sample persons less than 15 years old who move are not followed unless they move with a sample person 15 years or older, or attain their 15th birthday by the last day of the reference period in which they moved. In addition, persons who come into the survey after Wave 1 as a result of moving in with an original sample adult are not followed should they stop living with a sample person. Only original sample adults are followed. For more information about tracking persons over time, see Jean and McArthur (1986).

These rules are illustrated with a hypothetical example in figure 2.4. This example is used in subsequent chapters of the Guide to further illustrate various aspects of the sample design.

---

[8]When they return to the sample, these individuals are not treated like original sample adults in the event of subsequent moves, and their data cannot be linked over time.

[9]Note that unmarried children age 15 and over who enter college are not considered movers because residence in college dormitories is considered temporary. They will, instead, continue to appear as household members.

FIGURE 2.4 – **Illustration of the "Following Rules"**

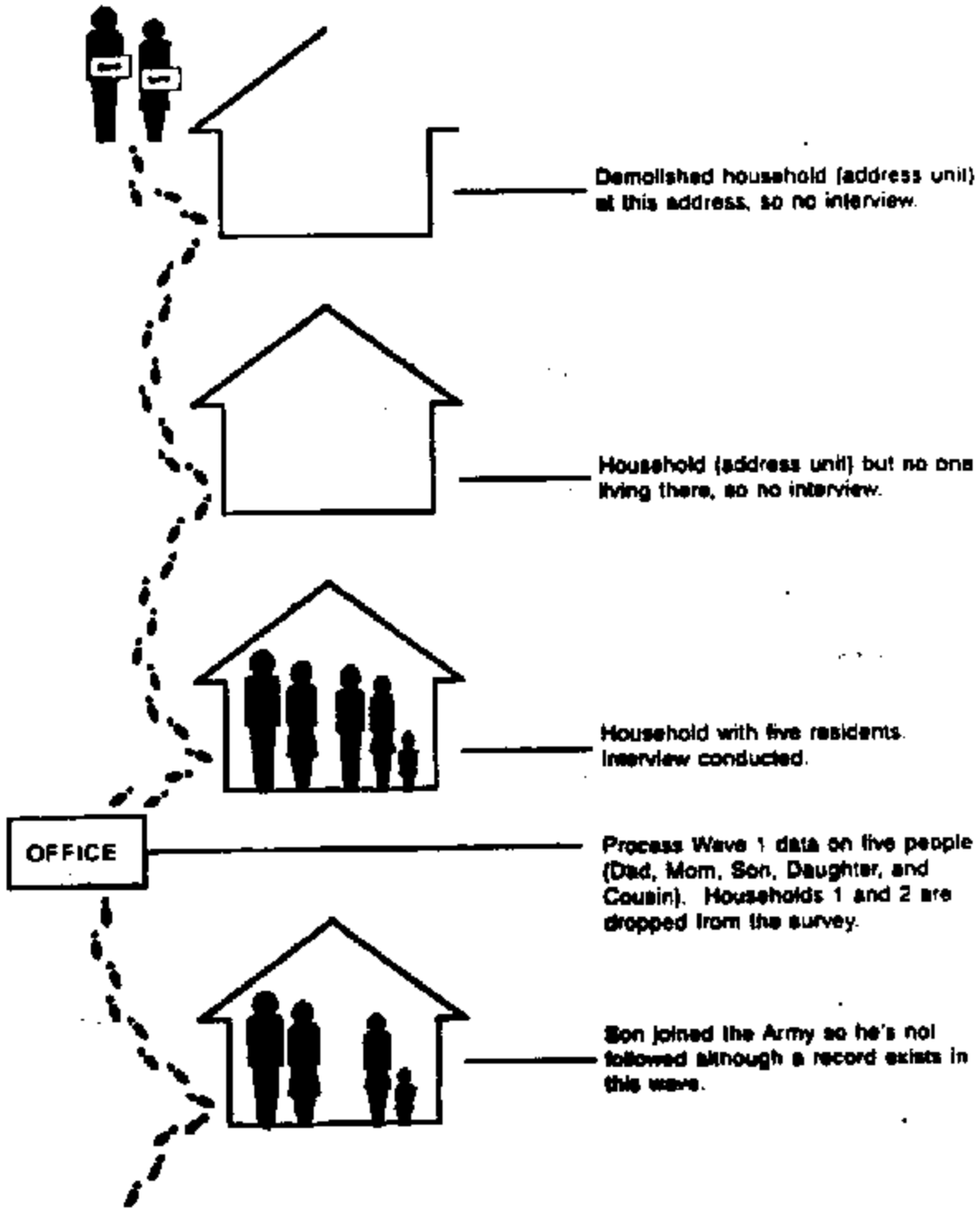


FIGURE 2.4 – **Continued**

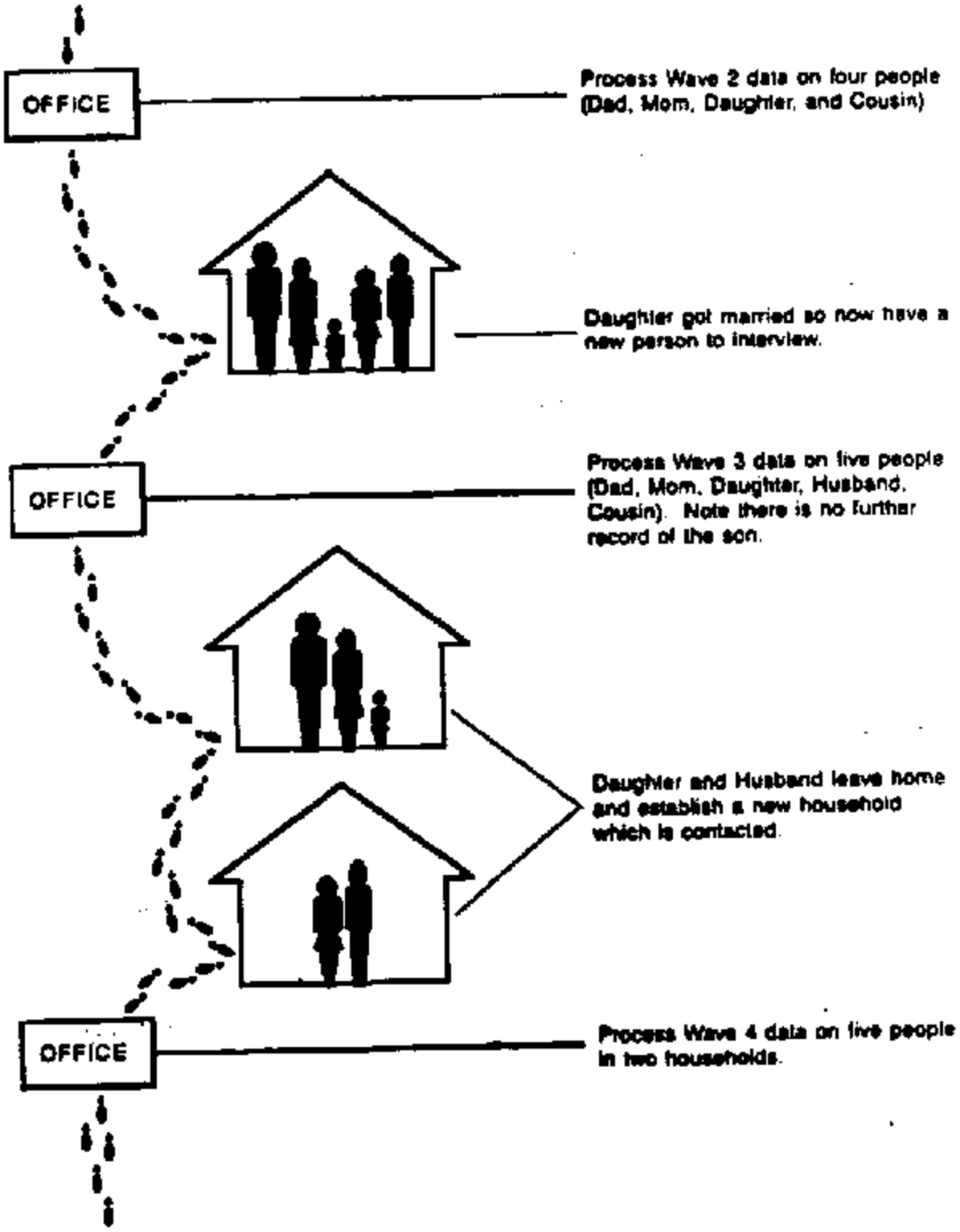


FIGURE 2-4 - **Continued**

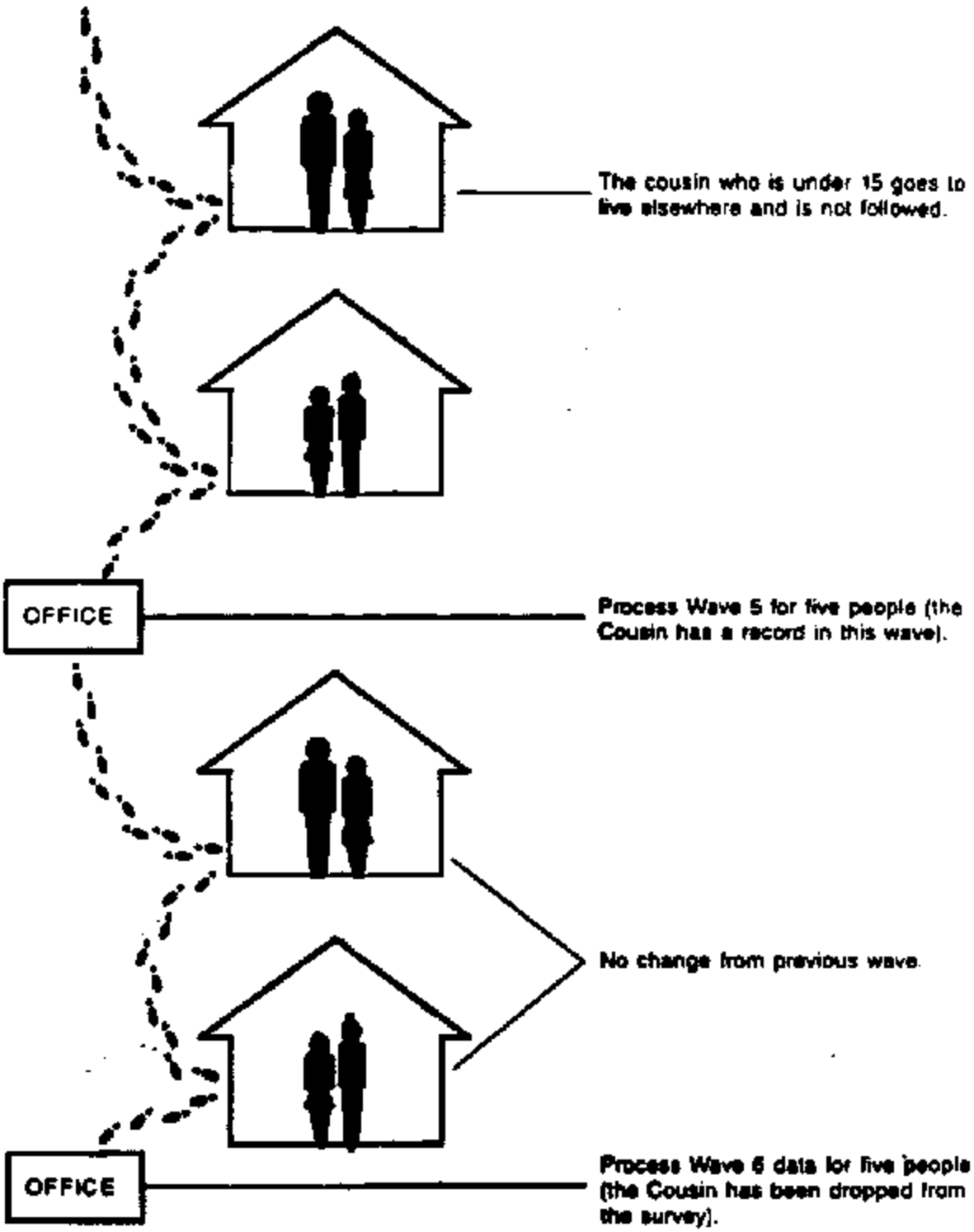


FIGURE 2.4 – **Continued**

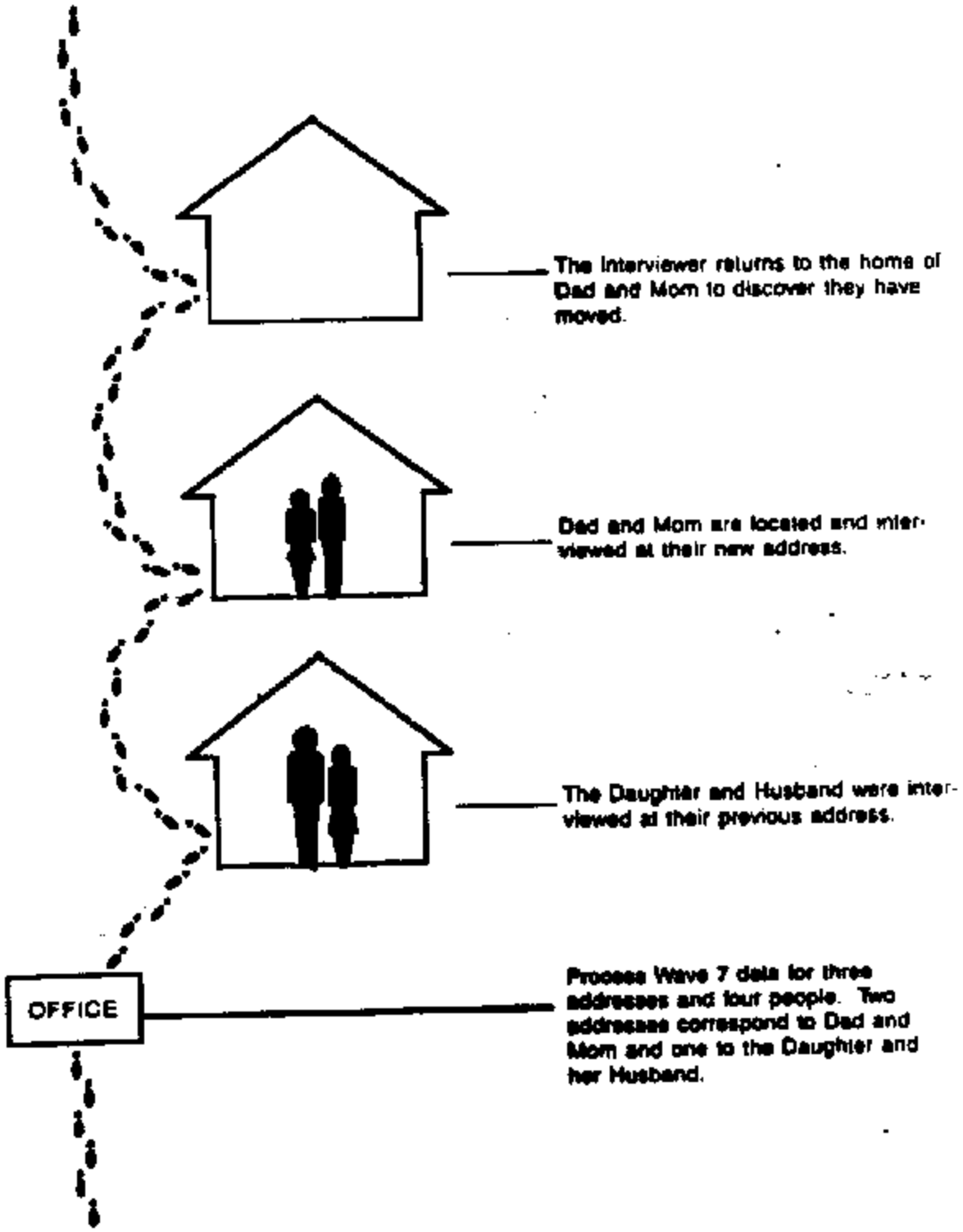
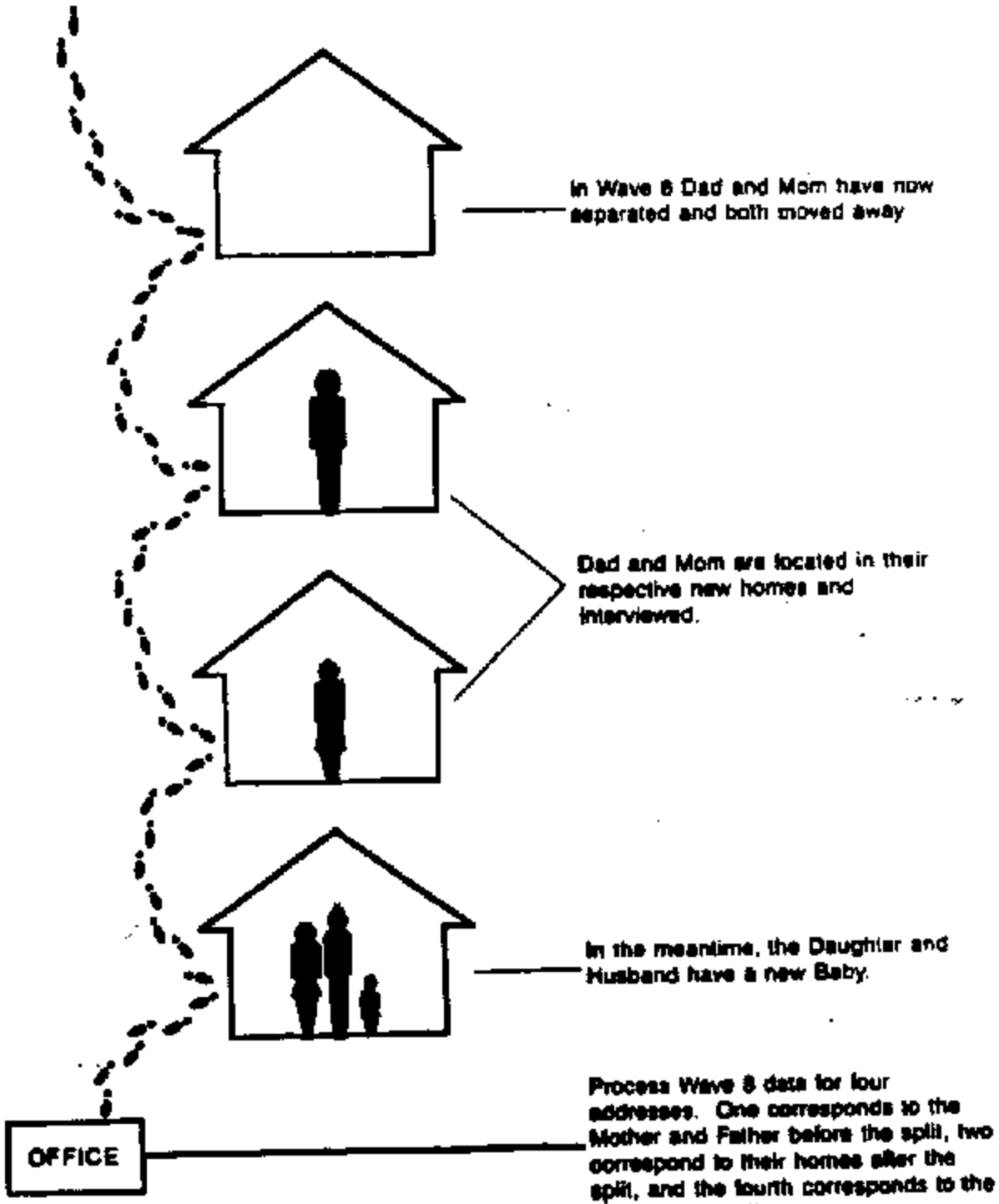


FIGURE 2 4 – **Continued**



REFERENCES

CITRO, C., D. HERNANDEZ, and R. HERRIOT. "Longitudinal Household Concepts in SIPP: Preliminary Results," Second Annual Research Conference Proceedings, March 23-26, 1986. Washington, D.C.: U.S. Bureau of the Census, 1986.

CITRO, C., D. HERNANDEZ, and J. MOORMAN. "Longitudinal Household Concepts in SIPP," Proceedings of Section on Social Statistics, American Statistical Association, 1986.

DUNCAN, G., and M. HILL. "Conceptions of Longitudinal Households: Fertile or Futile?" Journal of Economic and Social Measurement (Formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 361-375.

JEAN, A. and E. McARTHUR. "Tracking Persons Over Time," SIPP Working Paper Series No. 8701. Washington, D.C.: U.S. Bureau of the Census, 1986.

McMILLEN, D. and R. HERRIOT. "Toward a Longitudinal Definition of Households," Journal of Economic and Social Measurement (Formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 349-360.

CHAPTER 3 - SURVEY CONTENT

This chapter highlights the content of the Survey of Income and Program Participation (SIPP). While not all the variables are discussed in detail, the organization of the questionnaire is explained as well as its broad content. The chapter serves as a guide to what users may expect to find in the data files. The user should, however, obtain the complete questionnaire for each wave they wish to analyze because the content varies from wave to wave and from panel to panel. A SIPP Information Booklet, containing all questionnaires and flashcards used in the 1984 Panel, is available from Customer Services, Data User Services Division, Bureau of the Census, Washington, D.C. 20233 (301-763-4100). (Also, see Frankel, 1985.) The questions asked in the SIPP survey fall into two categories: those which require repeated measurement over the life of the survey and are included in every interview; and those asked only once or twice for each panel.

The first group of questions provides "core" data: labor force behavior, income, participation in public programs, basic demographic characteristics, and living arrangements. The frequency with which this information is collected should improve the measurement of income in particular. In addition, the design yields measurements for many different time periods and records patterns of change for each individual for longitudinal analysis.

In the second group of questions which are not asked repeatedly, the items may refer to previous work history or education, which need only be asked once, or to taxes paid, which are appropriately asked once each year. They may also cover subjects which require such detailed questioning that the Census Bureau cannot afford to include them more than once or twice. These questions, asked at only one or two interviews, are organized into "topical modules," which are asked after the first one or two interviews of each panel.[1]

Other selected questions, such as medicare claim number and social security number, are those asked only the first time an individual is interviewed in each panel. These are integrated with the core questions in the wave, but, in fact, are not subsequently repeated or updated.

Other aspects of survey content which are not immediately apparent from examination of the questionnaires are discussed elsewhere in this Users' Guide. For example, geographic information on microdata files is described in chapter 5.

DATA COLLECTION INSTRUMENTS

SIPP employs two data collection instruments, a control card and a questionnaire. The control card serves as a basic source of information for updating over successive interviews. For each address unit, there is one control card containing a record of basic demographic characteristics for each person residing at the address, as well as a limited amount of housing information. The control card is filled out during the first interview and updated as necessary in each subsequent interview.

---

[1] In the 1984 and 1985 Panels both Waves 1 and 2 contain core data only. Beginning with the 1986 Panel, only Wave 1 is restricted to core data.

Since the control card is associated with a specific address, a new control card is filled out for each new address resulting from movement of original sample members. A new control card is used when entire households move, as well as when new households are formed by any original sample person who moves to a new address.

The primary survey document is the questionnaire. In each wave a separate questionnaire is completed for every person 15 years old and over living with original sample members. The questionnaire itself is subdivided into modules for the core information. When topical data are collected, additional modules are added, the number of which varies depending on the number of topics administered in the wave.

Interviewers transcribe some information from one form to the other to facilitate matching of questionnaire data to control card data. Basic identifying information such as age, race, and sex are copied from the control card to each questionnaire. Since the control card also serves as a running record of certain items that help structure future interviews and monitor change, a few items of information (such as employment status and income sources) are transcribed from the questionnaire to the control card after the interview. These items can then be easily updated in the next interview, as well as serve as a guide to the interviewer in asking those questions appropriate to the respondent. After each wave of interviewing, both the control card and questionnaire are returned to the regional office for keying and processing. After keying and processing, the control card is returned to the interviewer to be used in subsequent interviews, while the questionnaires go to a central location for storage. When an address contains no original sample adult, the control card for that address is processed just after the last original adult leaves and is then archived.

INFORMATION ON THE CONTROL CARD

The control card contains administrative information, a few housing characteristics, basic demographic items for each person, and a variety of items to help keep track of people who move into or out of the address. A facsimile of the control card appears in the technical documentation and in the SIPP Information Booklet.

Each person at the address is listed on page 2 of the control card along with several demographic items: household relationship, date of birth, sex, marital status, race, and ethnic origin. Household relationship is expressed in terms of relationship to the "reference person," also known as the "householder." The reference person is the person in whose name the home is owned or rented. If the house is owned jointly by a married couple, either the husband or the wife may be listed as the reference person or householder. Family relationships among other household members are also recorded on the control card and, together with relationship to the reference person, are used to define families, subfamilies, unrelated subfamilies, and unrelated individuals. Family relationships data on the control card include relationship to reference person, person number of parent, marital status, person number of spouse, and designated parent or guardian.

If a person enters or leaves the household during the panel period (through birth, marriage, death, institutionalization, or migration), the date of the entry or exit is recorded so that most changes in household and family composition can be pinpointed to a specific reference month. Recording of entry and exit dates allows the computerized construction of households, families, and other relationships for each month covered by the survey. (The public-use microdata files carry five separate fields for items like marital status, relationship, and family type--one for each of the 4 reference months and a fifth for the interview date. They also carry flags denoting whether any change occurred in family or household composition within the 4-month reference period.)

Date of birth is recorded so that age can be computed for any subsequent interview or reference month. The public-use microdata files include year and month of birth, except that birth prior to 1900 is recoded to 1900 for confidentiality. Similarly, ages 85 or more are recoded to 85. Race is provided for four major race groups: White; Black; American Indian, Eskimo, and Aleut; and Asian and Pacific Islander. Ethnic origin is recorded in terms of 23 categories, including 7 categories of Spanish origin.

For persons under 15 years of age, the control card contains only the information on age, race, sex, household and family relationship, and entry and exit dates discussed above. The only other information collected about children in SIPP relates to their coverage by health insurance, food stamps and other programs, and their social security number. These data are obtained on the questionnaire of their parent or guardian who lives in the household.

For persons 15 years of age and over, the control card also includes information on education and veteran status. Two questions are asked concerning education: "What is the highest grade or year of school this person attended?" and "Did he/she complete that grade?" The response to these two questions allows calculation of the variable "Years of School Completed." (Current enrollment in school is not identified in the core questionnaire for the 1984 Panel, but is covered in topical modules in Waves 6 and 9. In subsequent panels, school enrollment has been added to the core.) Three questions are asked about military experience: "Did this person ever serve on active duty in the U.S. Armed Forces?"; if so, "When did this person serve?"; and "Is this person currently in the Armed Forces?"

Some limited data about the characteristics of the housing unit and the household are collected on page 1 of the control card. Information on number of units in structure, tenure (owning or renting), and size of household are noted. Other items identify whether the address is in a public housing project, or whether the occupants are receiving government rent subsidies.

One section of the control card is used for recording household noninterview information. Noninterview data are important in gauging the reliability of the survey and in calculating respondent weights. Household noninterviews are of four types, ranging from refusal of the household to be interviewed, to the interviewer finding the housing unit demolished. These types are detailed in chapter 7.

Other control card items are used to update the status of persons in future interviews. For example, after the interview, the interviewer transcribes the names of employers of all respondents aged 15 and over into item 42 of the

control card. Income sources and assets are also transcribed to the control card. These entries serve as references in conducting the next interview.

The control card also contains space for a telephone number for callbacks to verify information or obtain data unavailable at the interview (such as savings account balance). To help follow the household if they move before the next interview, the respondent is asked to furnish a telephone number of someone who would know their future address. Respondents' social security numbers are also entered in order to assist research projects that require merging administrative records with survey data. However, this information is not available in the public-use files.

CORE QUESTIONNAIRE

The questionnaire is modular in design, containing both items measured repeatedly and items which vary from wave to wave. The core, i.e., the part containing questions asked at every interview, includes five major sections:

1. Labor Force and Recipiency. Gathers information on labor force activity, whether the person participates in various programs, and the various types of income and assets the respondent receives or owns.

2. Earnings and Employment. Collects data on employment and earnings from specific wage or salary jobs, or self-employment.

3. Amounts of income received. Lists amounts of income from the sources identified in section 1 or from assets.

4. A short set of program questions. Asks this information only of the householder.

5. Missing wave. Collects previous wave data for persons who missed the preceding interview.[2]

A few observations will be useful in finding one's way through the questionnaire. Questions are numbered at the left starting at "1" within each of the sections listed above. A more important set of four-digit numbers, called source codes, appears in small boxes down the center of each page. These numbers are referenced in the data dictionary of the public-use microdata file and its index. (In fact, the index to the data dictionary can also be used as an index to the questionnaire.) These file source codes are consistent from wave to wave, even though the question numbers and page formatting of the questionnaire may change. These source codes are important aids to understanding the public-use tapes because the order of the questionnaire is incorporated into the data file, as described in chapter 4. The source codes are numbered according to the section of the questionnaire and appear in

[2]This module is not part of the Wave 1 core in any panel. It was also not included in the core for Waves 2 and 3 in the 1984 Panel. For a discussion of this section, the reason for its existence, analysis of data collected, and the decision to discontinue it in March 1987, see Huggins (1987).

different sections of the data files: 1000 series codes furnish information about the person; 2000, wage or salary income and self-employment income; 3000, items on program income; 4000, asset-based income and household program questions; and 6000, missing wave.

In examining any particular question, one needs to pay close attention to the sequence of prior questions as skip patterns are employed to restrict the universe for many questions. The skip patterns are complex, accounting for the differences among respondents. People who had no job, for example, should not be asked what type of job they had. In other situations, two questions may seem to ask for the same information, when, in fact, each is asked of different respondents. For instance, questions 2b and 7b in section 1 ask which weeks the respondent was looking for work or on layoff. Preceding questions, however, direct the interviewer to ask question 2b only of persons who did not work at all during the 4 reference months, while 7b is asked of persons who worked at least some during the period. Examining the sequence of prior questions in waves beyond Wave 1 is particularly important because a number of questions in the recipiency section, such as "reasons for receiving Social Security," are only asked of persons the first time they are interviewed.

Each section of the core questionnaire is reviewed below.

Labor Force and Recipiency

In section 1 of the questionnaire, "Labor Force and Recipiency," over 30 questions are asked of each respondent to determine their labor force status during the 4-month reference period. For persons who worked any time during the reference period, questions are asked on the number of weeks worked and weeks without a job. (The respondent is shown a calendar for the months and weeks of the reference period.)

For persons who were not working for all or part of the reference period, SIPP asks if they were on layoff or looked for a job. If they met these conditions, they are considered part of the labor force but unemployed. Those with no labor force activity, i.e., without a job, not on layoff, and not looking for work, are considered not in the labor force.

With these data one can examine labor force participation and how it changed during the 4-month period (see Ryscavage and Bregger, 1985). A person could be both employed and unemployed at different times during the reference period. Two characteristics--whether the respondent was working or looking for work--are recorded on a week-by-week basis. Part-time work (including the reasons for working part time) can also be studied using SIPP.

In computer processing many of the responses are recoded to make the data more manageable. For example, an employment status recode for each month of the reference period specifies whether the respondent had a job, worked at it full time, missed any work weeks, was laid off or looking for work, or had no job and was or was not looking for one. With other recodes, one could examine the number of weeks worked, the number of weeks unemployed, and the number of weeks looking for a job in each reference month.

Following the items on labor force, the focus of the questionnaire changes to the various governmental or other programs from which the respondent may receive income, as well as miscellaneous sources of income such as alimony, child support, and interest from savings. In this section the interviewer attempts to determine the list of income sources for which amounts were ever received in the 4-month reference period. Figure 3.1 shows the list of income sources used in the survey. Programs covered include private as well as government sources of income such as pensions, Social Security, Railroad Retirement, disability payments (e.g., Black lung and workmen's compensation), and survivor's benefits (e.g., life insurance). Means-tested government transfer programs include food stamps, Aid to Families with Dependent Children (AFDC), and other forms of public assistance for low income families. Miscellaneous sources such as income for foster child care are listed, as well as information about Medicare, Medicaid, and private health care coverage. Educational assistance such as Pell Grants and GI education benefits, student loans, and federally funded work-study programs are also covered. Note that in the public-use tapes, five of these program income sources are combined for confidentiality reasons (see figure 3.1).

Section 1 of the questionnaire also gathers information on ownership of the assets shown in figure 3.1 by each member of the household. As with unearned income recipiency, information on asset ownership is collected for the 4-month reference period preceding the interview month.

The items described in this section all have source codes in the 1000 series and thus appear in the data file as information on the record of each person. In addition to identifying income recipiency questions using the source code numbers, each unearned income type is assigned an Income Source Summary (ISS) (or income type) code. These codes, which are listed in figure 3.1, are frequently used in labeling the income amounts in the public-use microdata file and are used in defining the skip patterns for questions posed in "Amounts," section 3 of the questionnaire.

As income sources and assets are identified in this section of the questionnaire (various flashcards remind the respondent of possible income sources), they are posted to the ISS carried on the next-to-the-last page of the questionnaire. As noted above, each income type is identified using an ISS code. The ISS then serves as a reference for the interviewer in filling out other sections of the questionnaire, particularly section 3 where amounts of income received are obtained for each of the income sources identified. After the interview, the sources listed in the ISS are posted to the control card. At the next interview, the interviewer transfers this control card information to the new questionnaire. The interviewer then checks with the respondent to make sure he or she had actually received income from those sources or if any other sources were left out during the previous interview or not transcribed. Thus, the ISS and control card are used to continually update income information from wave to wave.

Earnings and Employment

While in section 1 general information is gathered about labor force activity, section 2 of the questionnaire asks the respondent about his or her specific type of employment and earnings. Slightly different questions are asked depending upon whether the person is a wage and salary employee and/or is self-employed.

FIGURE 3.1 - TYPES OF INCOME RECORDED IN SIPP

Wage or Salary Income

Income from Job #1
Income from Job #2

Self-Employment Income

Income from Business #1
Income from Business #2

Program and Miscellaneous Income (General Amounts Type 1)

ISS Code	Income Type
1	Social Security
2	U.S. Government Railroad Retirement pay
3	Federal Supplemental Security Income (SSI)
*4	State Supplemental Security Income (State administered SSI only)
5	State unemployment compensation
6	Supplemental Unemployment Benefits
7	Other unemployment compensation (Trade Adjustment Act benefits, strike pay, other)
8	Veterans compensation or pensions
*9	Black lung payments
10	Worker's compensation
*11	State temporary sickness or disability benefits
12	Employer or union temporary sickness policy
13	Payments from a sickness, accident, or disability insurance policy purchased on your own
20	Aid to Families with Dependent Children (AFDC, ADC)
21	General assistance or General relief
*22	Indian, Cuban, or Refugee Assistance
23	Foster child care payments
24	Other welfare
25	WIC (Women, Infants and Children Nutrition Program)
27	Food stamps
28	Child support payments
29	Alimony payments
30	Pension from company or union
31	Federal Civil Service or other Federal civilian employee pensions
32	U.S. Military retirement pay
*33	National Guard or Reserve Forces retirement
34	State government pensions
35	Local government pensions
36	Income from paid-up life insurance policies or annuities

*These five types of benefits are combined into a single category (#75) on the public-use microdata file to avoid increasing the risk of individual disclosure.

FIGURE 3.1 - CONTINUED

Program and Miscellaneous Income (General Amounts Type 1) (cont'd.)

ISS Code	Income Type
37	Estates and trusts
38	Other payments for retirement, disability or survivor
40	G.I. Bill/VEAP education benefits
50	Income assistance from a charitable group
51	Money from relatives or friends
52	Lump sum payments
53	Income from roomers or boarders
54	National Guard or Reserve pay
55	Incidental or casual earnings
56	Other cash income not included elsewhere
75	Five types (asterisked above) combined

Asset Income (General Amounts Type 2)

ISS Code	Income 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

Noncash Income (other than WIC and Food Stamps)

Medicare
Medicaid
Public housing occupancy
Rent subsidies
Energy assistance
Subsidized school lunches or breakfasts

The items in this section have source codes in the 2000 series for wage and salary employees and the 2200 series for the self-employed. These are grouped together on the public-use data files as described in chapter 4.

For wage and salary employees, data are collected on occupation, industry, and class of worker for up to two jobs. For each job, the employed respondent is asked to state the kind of business or industry in which he or she works, the kind of work done, and his or her main activities or duties on the job. These responses are later coded according to the Census of Population Industry and Occupation Classification Systems.[3] The detailed industry and occupation codes can be found in the technical documentation.

Several questions are asked to determine total pay from each job measured for each month of the reference period. Care is taken to avoid problems resulting from the fact that some people are paid on an hourly basis and some on a weekly or monthly basis. For persons paid on a weekly or a biweekly basis, questions are asked to determine the number of paychecks received that month. This helps pinpoint for the user those month-to-month income fluctuations that are merely due to variations in the number of weekly paydays each month.[4]

If a person worked on more than one job, the full information is collected for the two jobs worked the most hours. Earnings from other job(s) are combined into "other cash income not included elsewhere" (ISS code = 56).

For those respondents who are self-employed, the survey identifies the person's salary as well as other earnings from up to two businesses. As with wage and salary employees, data are collected on occupation and industry for each business, and these responses are later coded according to census industry and occupation classification systems. Questions about the business itself include whether the business was incorporated, a sole proprietorship, or a partnership; total value of the business; and net profits or losses for the year. Self-employment earnings in SIPP are defined by the income received each month by each person involved in the business. This concept, referred to as "draw," is different than the concept used in the March Annual Demographic Supplement to the CPS which defines self-employment earnings as net profit or loss.

Information is recorded separately for up to two businesses. As is true for wages and salaries, if income is received from more than two businesses in the reference period, the additional earnings are combined into "other cash income not included elsewhere" (ISS code = 56). The two businesses recorded are those with the most hours worked in the reference period.

---

[3]In the 1986 and later panels, dependent industry and occupation coding will be used subsequent to Wave 1. This means that when the reported employer or duties do not change from a prior wave, then the previous wave values for industry and occupation are assigned in the current wave.

[4]Income data are recorded as exact dollar amounts as reported by the respondent. Nonetheless, very high income amounts are top-coded on the public-use microdata files to avoid increasing the risk of individual disclosure. Chapter 5 contains a discussion of this "top-coding."

Amounts

Section 3 of the questionnaire is used to collect information on amounts of income received from each of the program and asset sources identified in section 1. A complete listing of these income sources is shown in figure 3.1. Section 3 consists of two main segments. Part A of section 3 deals with program and miscellaneous income, also known as "General Amounts Type 1," while those collected in parts B to F deal with income from assets, termed "General Amounts Type 2." General Amounts Type 1 variables have source codes in the 3000 series; General Amounts Type 2 source codes are in the 4000 series. As noted previously, each income type is assigned an ISS code for identification purposes. These ISS codes are listed in figure 3.1.

Though many different income sources are identified, the questionnaire only allows for up to six of the General Amounts Type 1 incomes to be recorded in the six repetitions of section 3, part A. In practice this is a negligible limitation since few people receive income from a larger number of covered sources.[5] Amounts of income from sources other than assets are collected for each month of the reference period for each adult reporting receipt in section 1. In collecting amounts of income from assets, items are carefully screened to eliminate double-counting of income received from property held jointly by husband and wife. Income from assets owned jointly by spouses is coded separately from income derived from individually owned assets. Asset income is collected for the 4-month reference period as a whole, while most other sources of income are reported month by month. The focus here is on income earned from each asset during the reference period, not on the value of the asset. For example, respondents are asked the interest earned from savings accounts; if that amount is provided, the balance in the accounts is not ascertained. Items are included to collect the total value of particular assets, but these are only administered when the amount of asset income is missing. The balances are not recorded in the public-use tapes.[6] Topical modules in two waves of each panel provide more complete information on asset ownership and asset values.

Unearned income other than asset income is recorded on the public-use tapes in monthly amounts for each source listed in figure 3.1.[7] Both the original reported monthly amounts and imputed amounts are retained in the files, as are

---

[5]Also included on the public-use files in the person section is a list of income sources for which recipiency was reported in the reference period. The variable INCSOURC refers to unearned income other than that received from assets, and the variable ASTSOURC refers to asset income. Individuals reporting more than 6 unearned income types will have the complete list (up to 10) included in the first variable.

[6]There are two exceptions to this rule. In each of the two waves of a panel where asset values are the subject of the topical module, the questions on asset values are administered to everyone owning the asset. The responses are included in the public-use tapes as part of the topical module.

[7]Sources flagged with asterisks in figure 3.1 have been combined into one source (ISS code = 75) for confidentiality reasons.

flags to denote when the amounts were imputed. Asset income is also recorded as the original (lump sum) reported amount and recorded in monthly edited and imputed amounts. To create the monthly variables, the lump sum amounts after imputation are divided by four, and amounts received from joint accounts are evenly allocated to both spouses. Original reported and imputed amounts which are not evenly divisible by four have the remainder allocated in whole dollar amounts, starting in the earliest reference month until the remainder is exhausted.[8]

Program Questions and Noncash Income

The last section of the core questionnaire concerns participation of the household in selected programs. The questions in this section are asked of only the reference person at each address and recorded on the household record of the public-use files. These items determine whether the household participates in energy assistance and/or subsidized school lunch or breakfast programs. In Wave 1 of each panel, this section also measures benefits from subsidized housing programs. This is usually section 4 of the questionnaire, but in some waves (for example, Waves 4 and 7 of the 1984 Panel) it follows the topical modules as section 5.

Missing Wave

In Wave 4 of the 1984 Panel, a new section was added to the core questionnaire, called the "Missing Wave." The Missing Wave does not collect information relevant to the current reference period; rather it is used to collect selected information for original sample members at Wave (n+1) who were interviewed in Wave (n-1), but not interviewed in Wave (n). These respondents are asked about their employment; any weeks they may have been on layoff; and, if without a job for any period of time, whether they looked for work. Information is collected on income sources they had during the previous period that they did not still have in the present period, and on sources this period that they did not have in the previous period. Months in which the income was received are also recorded. Collecting this retrospective information will reduce the amount of imputation when constructing longitudinal files from SIPP. The Missing Wave appears in subsequent waves through the remainder of the 1984 Panel.[9]

TOPICAL MODULES

In addition to the core data collected every 4 months, most interviews for each panel include additional sets of questions called topical modules. The topics covered in these modules do not require repeated measurement during the year and may use a longer reference period than the 4-month period used for the core questions. (A topical module is termed "fixed" if the Census Bureau intends to conduct it periodically; it is termed "variable" if there is no schedule to periodically conduct the module.)

---

[8]Individuals reporting less than $4 for the reference period will therefore appear to have received the income only in the early months.

[9]Huggins (1987) evaluated the data collected in the missing wave section of the questionnaire and recommended that its use be discontinued. Effective March 1987, interviewers were instructed to stop using the form.

Topical modules are a distinguishing feature of SIPP. Their purpose is to address significant program and policy questions that do not require updating on each wave. Topical modules also give SIPP the flexibility to accommodate topics on emerging issues on relatively short notice. SIPP was designed to permit linkage of data across waves. Topical module as well as core data can be linked for individual persons, making possible the compilation of a number of topical variables that may bear on a particular research question. Such an array of information can be applied, for example, to eligibility criteria for social support programs. The extent of nonparticipation by eligible persons, or participation of noneligibles, could be measured. Or program analysts could simulate a change in criteria to study the possible impact on program participation (Klein, 1983).

The list of topical modules administered to each panel and the waves in which they are administered varies from one panel to the next. As an illustration of the types of information collected, the module topics and collection periods for the 1984, 1985, 1986, and 1987 panels are shown in figures 3.2, 3.3, 3.4, and 3.5. These modules are assigned to particular waves and may be repeated in later waves if more than one observation is necessary. For example, a wealth module is administered in Waves 4 and 7 of the 1984 Panel, during the same months one year apart. The detailed data on personal and household assets and liabilities collected for this module will allow an examination of economic well-being beyond that which can be studied with income data alone.

An annual "round-up" module is administered in the waves after the end of the first and second years of interviewing (Waves 6 and 9 for 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. This module also includes information on taxes, employer-provided benefits, and educational enrollment.

Other topics, such as health and disability, are administered in only one wave of the panel.

Topical modules are asked only of persons 15 years old and over and coverage varies depending on the topic. For example, only persons 25 or over are asked the questions on retirement and pensions (Wave 4, 1984 Panel). Questions on shelter costs and energy usage are asked only of the householder. Questions are asked about characteristics or coverage of children in only a few cases. For example, in the topical module on health and disability in Wave 3, 1984 Panel, parents are asked about disabilities of their children.

The processing system was developed to expedite the publication of core data. Topical module information is reviewed and edited in separate operations, and reports and microdata files including topical data generally appear later than those with core data collected in the same interview. (Data processing by the Census Bureau will be the topic of subsequent editions of the Users' Guide.)

The topical modules are discussed below.

Health and Disability

In the Health and Disability Topical Module, people are asked questions about the general condition of their health, functional limitations, work disability, and the need for personal assistance. Also included are hospital stays or

FIGURE 3.2 - TOPICAL MODULES FOR THE 1984 PANEL

	1983-1984		1985		1986	
Int. Month	Wave	Topical Module	Wave	Topical Module	Wave	Topical Module
Jan-Apr	1	(Oct-Jan) No topical modules	5	Support for Non-Household Members Reasons for Not Working/ Reservation Wage Child Care Arrangements and Expenses Welfare History Work-Related Expenses	8	Support for Non-Household Members Household Relationships Marital History Fertility History Migration History (Jan-Mar)
	2	(Feb-Apr) No topical modules				
May-Aug	3	Health and Disability Education and Work History	6	Annual Income Taxes Employee Benefits Educational Financing and Enrollment Training Questions	9	Annual Income Taxes Employee Benefits Educational Financing and Enrollment Training Questions (Apr-Jul)
Sept-Dec	4	Assets and Liabilities Pension Plan Coverage Job from Which Retired Retirement Plans Shelter Costs and Energy Usage	7	Assets and Liabilities Pension Plan Coverage		

FIGURE 3.3 - TOPICAL MODULES FOR THE 1985 PANEL

Interview Dates	Wave	Fixed Topical Module	Variable Topical Module
Feb 85- May 85	1	None	None
June 85- Aug 85	2	None	
Sept 85- Dec 85	3	Assets Liabilities	
Jan 86- Apr 86	4	Marital History Fertility History Migration History	Household Relationships Support for Non-Household Members Work Related Expenses
May 86- Aug 86	5	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	
Sept 86- Dec 86	6		Child Care Arrangements Child Support Agreements Support for Non-Household Members Job Offers Health Status and Utilization of Health Care Services Long-Term Care Disability Status of Children
Jan 87- Apr 87	7	Assets Liabilities	Pension Plan Coverage Lump Sum Distributions from Pension Plans Characteristics of Job from which Retired Characteristics of Home Financing Arrangements
May 87- Aug 87	8	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	

FIGURE 3.4 - TOPICAL MODULES FOR THE 1986 PANEL

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

FIGURE 3.5 - TOPICAL MODULES FOR THE 1987 PANEL *

Interview Dates	Wave	Fixed Topical Module	Variable Topical Module
Feb 87- May 87	1	None	None
June 87- Sept 87	2	Fertility History Marital History Migration History Recipiency History Employment History Work Disability History Education and Training History Family Background Household Relationships	Personal History
Oct 87- Jan 88	3		Child Care Arrangements Child Support Agreements Support for Nonhousehold Members Work Related Expenses Housing Costs Energy Usage
Feb 88- May 88	4	Assets Liabilities	Pension Plan Coverage Lump Sum Distributions from Pension Plans Characteristics of Job from which Retires Characteristics of Home Financing Arrangements
June 88- Sept 88	5	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	
Oct 88- Jan 89	6		To be determined in 1987
Feb 89- May 89	7	Assets Liabilities	
June 89- Sept 89	8	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	

*The short wave (interviews of three out of four rotation groups) was eliminated beginning with the 1987 Panel.

periods of illness, health facilities used, and health insurance plans (private or Medicare) available. If their children have a physical, mental, or emotional problem, the survey asks what caused the problem and whether they attend regular school. This module is included in SIPP because health and disability status are among the major factors affecting a person's work, earnings, income sources, and participation in public programs. (See Current Population Reports, Series P-70, No. 8.)

Education and Work History

For the Education and Work History Topical Module, people are asked questions about education such as their highest level of schooling attained; courses or programs studied in high school and beyond; whether they received job training; and, if they received training, for how long and under what program (e.g., CETA or WIN). They are also asked questions about their job history, including description of selected previous jobs and their rate of pay. Respondents who are not working are asked about their last job that lasted at least 2 weeks, or why they have never worked for such a period. Questions also refer, if appropriate, to when the respondent first worked 6 consecutive months at one job or business. The module then goes on to record the person's general work history and reasons for periods spent not working. The total module provides a basis for understanding a person's earnings income in conjunction with educational background. This module was incorporated into a series of "history" modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

Assets and Liabilities

The Assets and Liabilities Topical Module is included in SIPP to broaden our understanding of resources available to households and individuals. These data allow an examination of economic well-being beyond that which can be observed through the study of current income. Participation in many Federal programs is contingent upon not just the basic income level, but upon the assets held by the individual or household as well. Some of the major types of assets covered by this topical module are savings accounts, stocks, mutual funds, bonds, Keogh and IRA accounts. Unsecured liabilities such as loans, credit cards, and medical bills are also covered. Value of home and automobile are also collected in another topical module of this wave.

Some information on assets and liabilities is gathered in the core of the questionnaire, within the "Earnings and Employment" and "Amounts" sections. The topical module questions ask for the value of certain assets as of the last day of the reference period, such as the balance in checking accounts, the value of U.S. savings bonds, or the amount in an IRA account. Respondents are also asked about outstanding debts and obligations, such as unpaid bank loans or credit card bills. Debts owed jointly are distinguished from those owed by an individual. (See Current Population Reports, Series P-70, No. 7.)

The Assets and Liabilities Topical Module is repeated twice per panel, one year apart. In Wave 7 of the 1984 Panel, this module was used to examine two methods of measuring wealth and savings estimates. Asset and liability information collected in Wave 4, 1984 Panel, was provided to one-half the respondents interviewed in Wave 7. These respondents then had a reference point for reporting

current values. Their response patterns are being evaluated in comparison to those not given the feedback information to assess improvements in reliability and consistency, as well as possible biases introduced by the feedback process. (For example, some respondents may tend to report no change in order to shorten the interview.) Further comparisons are possible with the 1985 Panel-Wave 3, which included the same module administered to the respondents in the 1984 Panel who were not provided feedback of their earlier responses (Lamas and McNeil, 198[illegible].

The Retirement and Pension Topical Module contains questions on coverage and vested rights in retirement or pension plans. People are asked such questions as when they expect to stop working, whether they will eventually receive retirement benefits, whether their employer has a retirement plan and whether they are included in the plan, and how much they expect to receive per year from these plans. This topical module will provide information relating to retirement decisions.

Housing Costs and Energy Usage

The Housing Costs, Conditions, and Energy Usage Topical Module collects information on mortgage payments, real estate taxes, fire insurance, principal owed, when the mortgage was obtained, and interest rates; rent; type of fuel used and heating facilities; appliances; and vehicles.[10] Questions on value of home and automobile will be used in conjunction with assets and liabilities reported in the Assets and Liabilities Topical Module in order to calculate each individual's net worth. This topical module will also help fulfill a need for information concerning energy usage due to the increased interest in recent years over the rising costs of energy and concerns about conservation. The information can be used in analysis of the requirements of individuals and households who participate in energy assistance programs.

Child Care

The Child Care Topical Module includes questions 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. If the designated parent or guardian of children under 15 years of age in the household worked or went to school during the reference period, he or she is asked to furnish child care information for up to as many as three youngest children for the last reference month. The module asks whether a relative (including the other parent) or nonrelative cared for the child, if the child was in school, or if the child cared for self. The location of child care is asked and, if outside the child's home, whether the facility was licensed. Also collected is information on who provided transportation to the facility and hours per week of child care. The same information is collected for both primary and secondary child care arrangements, if any. Respondents are asked to provide data on the cost of child care, and whether he or she lost any time from work or school due to child care problems. These types of information are useful in the analysis of

[10]Subsequent to the 1984 Panel, questions on energy usage were combined into a separate module. Vehicles and housing values are retained together in a module entitled "Real Estate and Vehicles."

labor force behavior. Also, child care costs are a major part of work-related expenses and are frequently deductible from income in testing for government program eligibility purposes. (See Current Population Reports, Series P-70, No. 9.)

Welfare History

The questions in the Welfare History and Child Support Topical Module will help determine how long persons may have received aid from specific welfare programs and obtain information on child support agreements and their fulfillment. The data from the welfare history questions will measure the extent to which persons and households have been dependent upon government transfer programs in their general finances and will help evaluate the effectiveness of the programs.

One series of questions in the module concerns the food stamp, AFDC, and SSI programs. Current recipients are asked how long they have been receiving, or have been authorized to receive, these benefits. Recipients and nonrecipients are asked whether they had at any previous time applied for benefits, whether they received them and, if so, when and for how long. This module was incorporated into a series of history modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

The Child Support Topical Module attempts to determine whether those entitled to receive it have in fact received child support payments. The module asks whether the child support agreement was court-ordered or arranged otherwise and how the payments were to be made. It also asks for the amount and regularity of payment and whether a child support enforcement office has provided any help.

Reasons for Not Working

The Reasons for Not Working/Reservation Wage Topical Module ascertains 1) the reasons why persons are not in the labor force, and 2) the conditions under which persons might want to join the labor force. The reservation wage questions ask about the pay rate that a person would require in order to begin working (Ryscavage, 1987). Questions were also asked about job search and, if people have been offered but did not accept a job, the reason they refused it. This module was discontinued after the 1985 Panel.

Support for Nonhousehold Members

The Support for Nonhousehold Members/Work-Related Expenses Topical Module will aid in obtaining a measure of the fixed financial obligations of persons in order to obtain a better understanding of their economic situations. It contains questions that ask people about whether they provide regular payments to support persons who are not members of their household. It also asks about expenses associated with their job such as union dues, licenses, permits, special tools, uniforms, or travel expenses.

Annual Earnings and Benefits

The Annual Earnings and Benefits Topical Module includes questions that ask people about their calendar-year wages and salaries and income from their own business, as well as the receipt of certain employer-provided benefits not

covered elsewhere in SIPP, such as the use of a company car or truck, an expense account, or provision of free meals and lodging. In addition, a series of questions is administered about reasons for leaving for those persons who left a job during the calendar year. Questions about calendar-year earnings, taxes, health and life insurance deductions, and retirement contributions are designed to obtain the most accurate data available, and respondents are encouraged to refer to W-2 forms and other records. This module is administered twice per panel.

Property Income and Taxes

In the Property Income and Taxes Topical Module, people are asked for information on rental income received during the calendar year and on interest earned and/or dividends from assets such as savings accounts, money market deposit accounts, interest-earning checking accounts, bonds, or stocks. They are also asked about Federal and state income tax liabilities and certain other tax information such as type of return, use of selected schedules (for example, Schedule A, Itemized Deductions; Schedule B, Interest or Dividends; or Form 4835, Farm Rental Income), and number of exemptions. The tax questions are being asked in order to better estimate the distribution of after-tax income and to help build better microsimulation models of the tax and transfer system. This module is administered twice per panel.

Education Financing and Enrollment

The School Enrollment and Financing Topical Module contains a brief set of questions about schooling costs and sources and amounts of educational assistance. Asked of persons 15 years old and over, the module includes questions to pinpoint the grade level of persons enrolled in a general, technical, or business school; amount of tuition and fees; costs of room and board; and books and supplies. Specific sources of educational assistance, such as the GI Bill or employer assistance, are asked.

The questions in the Job-Related Training Topical Module refer to work history training programs. Persons 15 to 64 years of age are asked whether they have received job training, where they received it (apprenticeship, school, military, and so on), the duration and cost of the training, how it was paid for, and whether the training was federally sponsored.

Household Relationships

In the Household Relationships Topical Module, exact relationships are pinpointed for households of three or more members. This goes beyond the information on the control card described earlier in this chapter, which establishes relationship to the reference person only. The module establishes a matrix of relationships of each household member to all other members. Interviewers show the respondent, who must be the reference person, a flashcard defining the various possible types of relationships. Relationships are specified in detail. For example, a brother is identified as full brother, half brother, stepbrother, or adoptive brother. In-law relationships are also identified.

Migration History

The Migration History Topical Module asks respondents 15 years of age and older when they moved to their current residence, the location of any previous residence and when the respondent lived there, and their state or country of birth. Respondents born in a foreign country are also asked about their citizenship status and when they came to the United States to stay. During some, but not all years, additional questions are also asked, including state or country of birth of both mother and father, reasons for moving to current residence, several questions on payment of moving expenses, and several questions on current and second residences. The Migration History Topical Module was incorporated into a series of history modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

Marital History

The Marital History Topical Module questions are asked of all respondents 15 years or older who have ever been married. Date of present marriage is determined; for those married more than once, the dates of their first two marriages. If appropriate, respondents are asked when their previous marriages ended, and whether they were widowed or divorced. The number of times married is also asked. The Marital History Topical Module has been incorporated into a series of history modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

Fertility History

Fertility History Topical Module questions are asked of females 15 years old or more and males 18 or older. Men are only asked the number of children they have fathered. Women are also asked about their birth expectations in the future. Women who have ever had children are asked when their first and last children were born and, if born 1960 or later, where they now live. Mothers of children born 1960 or later are asked about their employment status prior to the birth and during pregnancy with the first child. If employed during that pregnancy, they are asked about the circumstances of absence from the job before and after the birth and the maternity leave policies of their employers. Post-birth employment is also covered. The Fertility History Topical Module was incorporated into a series of history modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

Personal History

For the 1986 and later panels, the Census Bureau developed the Personal History Topical Module to address major policy and program concerns. The questions in the topical module will help reduce, if not eliminate, the "left-censoring" analysis problem caused by not knowing when current situations covered by the sections described below began. This module incorporates questions from a variety of topical modules described previously. The Personal History Topical Module contains the following modules.

The Recipiency History Topical Module will help determine how long persons may have received aid from selected programs. Data from these questions will measure the extent to which persons and households have depended upon government transfer programs and will help evaluate the effectiveness of the programs.

The Employment History Topical Module supports analysis of persons' past labor force patterns and the relationship to their current employment status and their degree of reliance on government programs.

The Work Disability History Topical Module determines how a person's health and disability will affect his or her work, earnings, and participation in public programs.

The questions in the Education and Training History Topical Module will provide a basis for understanding persons' current labor force status, based on their education and training background.

Persons in the age range 25-64 respond to the questions in the Family Background Topical Module to obtain an indication of the respondent's family characteristics at the time of his or her 16th birthday. Such characteristics include how many brothers and sisters the person had, who the person lived with when he or she was 16 (both parents, father only, etc.), and the highest grade completed and occupations of the parent(s).

The questions in the Migration History Topical Module obtain place of birth, duration of stay at current residence, and the time of the latest interstate move, if any.

The Fertility History Topical Module obtains information on birthdate of up to three children. These data can be used in conjunction with the Employment History Topical Module to examine labor force changes related to the birth of children.

The Marital History Topical Module obtains marriage, divorce, widowhood, and separation data for all household members 15 years of age and older who have ever been married. This module will obtain data for up to three marriages -- the first two and the most recent. For persons married more than three times, marriages between the second and most recent are not covered in this module.

The Household Relationships Topical Module determines the family and nonfamily relationships that link each person in the household. Interviews will include this module only once per household, on the reference persons' questionnaire.

REFERENCES

FRANKEL, D. "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, 1985.

HUGGINS, V. "Evaluation of Missing Wave Data From the Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, forthcoming.

Klein, Robert E. "The Feasibility of Topical Modules as a Design Feature of Panel Surveys: Lessons From the 1979 Income Survey Development Research Panel," Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program, ed. Martin H. David. New York: Social Science Research Council, 1983.

LAMAS, E. and J. McNEIL. "An Analysis of the SIPP Asset and Liability Feedback Experiment," Proceedings of the Section on Social Statistics, American Statistical Association, Forthcoming.

RYSCAVAGE, P. "An Evaluation and Analysis of Reservation Wage Data for SIPP," Proceedings of the Section on Social Statistics, American Statistical Association, Forthcoming.

RYSCAVAGE, P. and J. Bregger. "New Household Survey and the CPS: A Look At Labor Force Differences," Monthly Labor Review, September, 1985.

U.S. Bureau of the Census. "Household Wealth and Asset Ownership: 1984," Current Population Reports, Series P-70, No. 7. Washington, D.C.: U.S. Government Printing Office, 1986.

U.S. Bureau of the Census. "Disability, Functional Limitation, and Health Insurance Coverage: 1984/85," Current Population Reports, Series P-70, No. 8. Washington, D.C.: U.S. Government Printing Office, 1986.

U.S. Bureau of the Census. "Who's minding the kids? Child Care Arrangements: Winter 1984-85," Current Population Reports, Series P-70, No. 9. Washington, D.C.: U.S. Government Printing Office, 1987.

CHAPTER 4 - STRUCTURE OF CROSS-SECTIONAL PUBLIC-USE MICRODATA FILES

The Survey of Income and Program Participation (SIPP) public-use microdata files consist of unaggregated records for individual survey respondents, edited to remove information that might be used to determine the identity of any respondent. Microdata files allow a user the flexibility to perform a wide variety of analyses, perform statistical tests, and produce different tabulations of the population from those published by the Census Bureau.

The microdata are oriented to three subject areas:

- o "Cross-sectional core" files, each containing core labor force, income, and program participation data from one wave of interviews;
- o "Topical module" files, containing both core information and cross-sectional data from topical modules asked at one interview; and
- o "Longitudinal" files containing core data from multiple interviews matched together for each respondent.

This chapter discusses the cross-sectional core files. (Material on the other files will be included in updates to the Users' Guide.) The cross-sectional core files may also be called wave files, as each contains data from one wave of interviewing (see chapter 2, "Survey Design"). The Census Bureau issues these files in two formats, complex (also referred to as relational) and rectangular. The chapter, therefore, opens with a section explaining these file structures and the relative advantages of each to the user.

The files are then discussed in turn. The complex file contains several record types. These are described, with a brief summary of content in each. This section also explains how the user associates these records with persons, families, and households. An example is provided. The sort sequence of records is also explained. Users interested in this file should also read the section on the rectangular file for further clarification of concepts common to both files.

The chapter then describes how the rectangular version of the data file is structured. Those interested in this file structure should also read the section on the complex file, as information common to both files is not repeated in the rectangular file discussion. The rectangular file section notes certain recodes found only on this version of the file and explains the arrangement of data items on each record. The sort sequence of the file as a whole is also explained, and use of the file is illustrated by example.

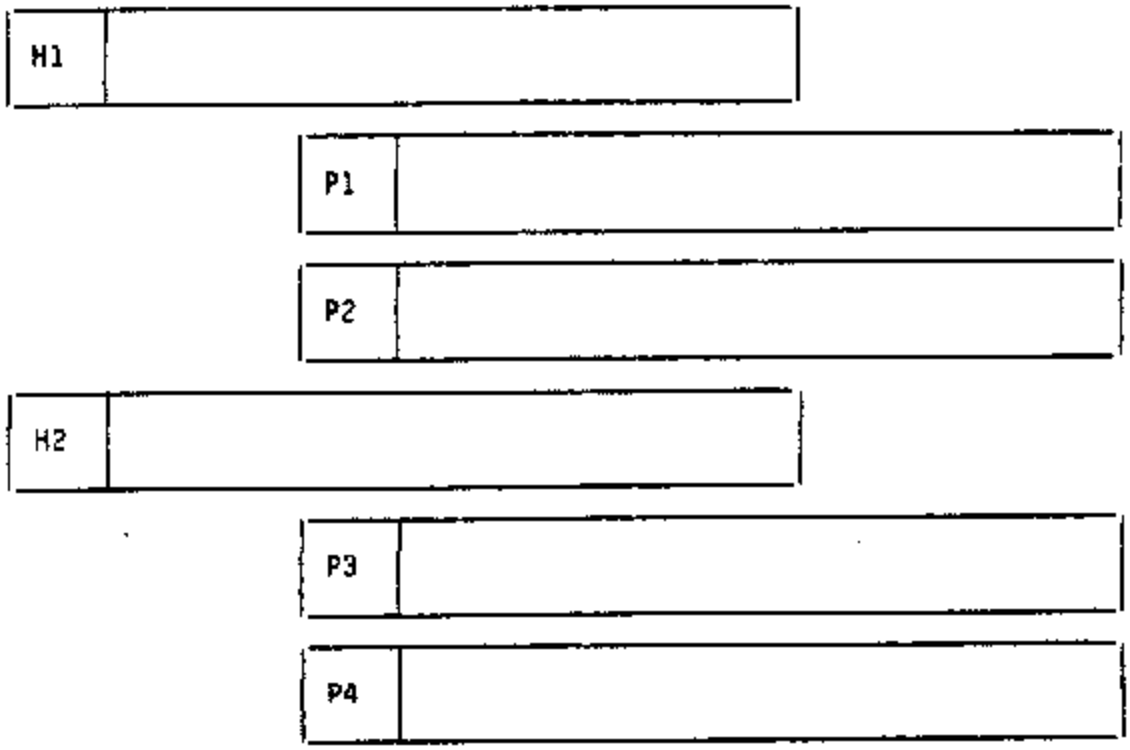
The chapter concludes with sections on reading the data dictionary; technical specifications and options for ordering data tapes; and references to other chapters in the Users' Guide on estimation, accuracy of the data, and matching files from more than one wave of interviewing.

ALTERNATE FILE STRUCTURES

Most current Census Bureau demographic surveys, such as the Current Population Survey (CPS), as well as the public-use microdata samples from the decennial

censuses, are released as hierarchical files. A hierarchical file consists of more than one type of record, with the records at one level grouped together with the corresponding records at other levels. For instance, figure 4.1 represents a simple hierarchy of households and persons within them. The first two persons (P1 and P2) are members of household 1, while the third and fourth persons (P3 and P4) are members of the second household. On the data file, the household record, with information on the household as a whole, appears first, followed by separate person records for each household member.

FIGURE 4.1 - HIERARCHICAL FILE STRUCTURE



At first, SIPP seems amenable to a hierarchical structure. There are addresses, within sample units; families are within addresses; persons are within families; and income sources are associated with persons. Because of the richness of SIPP data, however, that structure does not work. Each file has 5 reference months--the interview date and the 4 prior months--and household and family composition and personal relationships are allowed to change from one month to the next. If, for example, a daughter and son-in-law leave the home of the daughter's parents and form a new household in the third month of the 4-month reference period, a different set of hierarchical relationships exist over time. One household, containing two family units, exists in months 1 and 2 and two households, each with one family unit, exist in months 3 and 4.

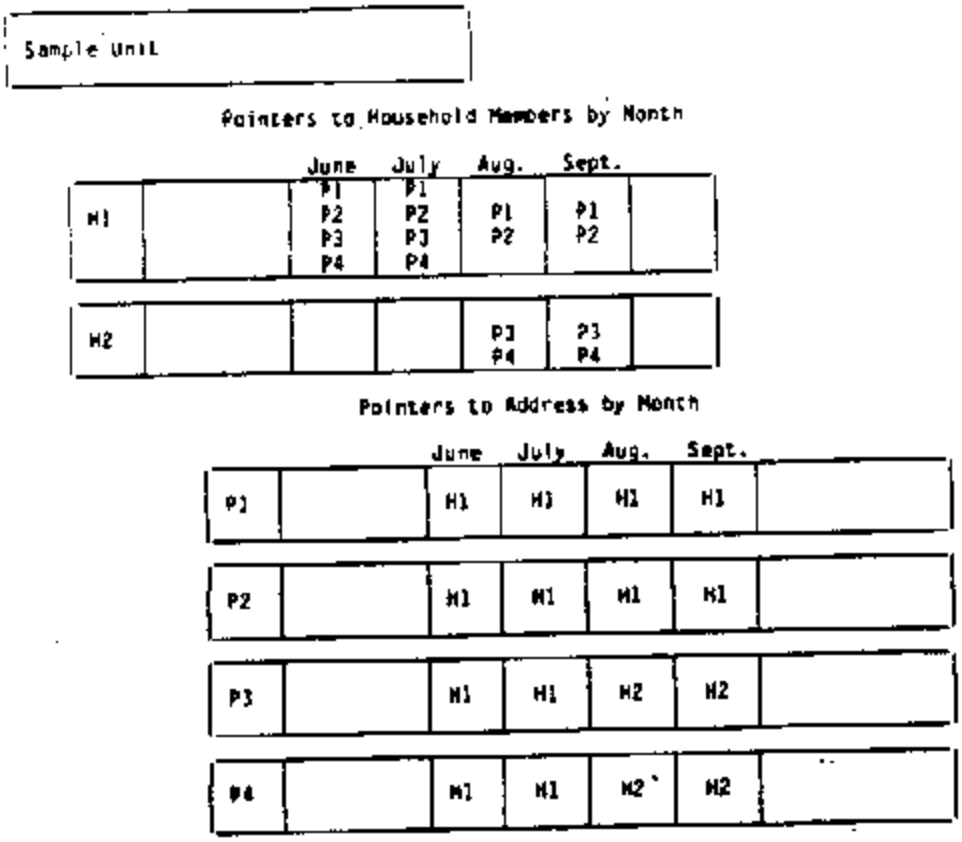
The user should be aware of the full implication of the measurement of monthly household and family composition in using the SIPP data products described in

this chapter. Because composition is not fixed, there is no longitudinal concept which can be used to link the monthly household and family units over time. Users desiring to learn more about the difficulties in defining longitudinal units are referred to McMillen and Herriot (1984); Duncan and Hill (1985); Citro, Hernandez, and Herriot (1986); and Citro, Hernandez, and Moorman (1987).

In order to accommodate changing composition within a wave, the SIPP files for each wave must structure the information in a flexible way. A hierarchical file defines relationships implicitly by the relative position of records. The user knows that persons P1 and P2 in figure 4.1 belong to household H1 because that is the record they follow. But in the Census Bureau's internal SIPP data base, these relationships are expressed by pointers and the use of these pointers has been carried over to the SIPP public-use files. Records P1 and P2 in figure 4.2 indicate with explicit pointers that they had a common address for all 4 months, while records P3 and P4 moved from one address to another. Correspondingly, there are pointers on the address records to the person records of their members for each month.

An advantage of using pointers becomes evident when there are many relationships to express. Consider the example, illustrated in figure 4.2, in which the daughter and son-in-law (P3 and P4) move out of her parents' household (H1) at the beginning of August. The June and July pointers link them with address H1, but the August and September pointers show them as residing at address H2.

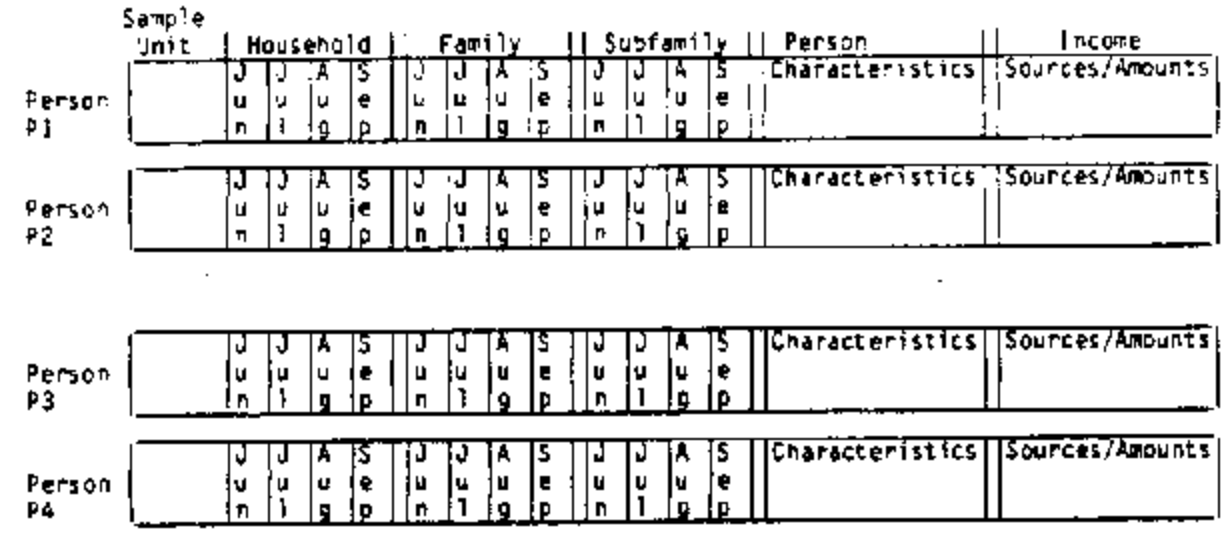
FIGURE 4.2 - MONTH-TO-MONTH CHANGES SHOWN IN A COMPLEX FILE



Of course, the structure of SIPP is much more complex than in this example. The example ignores, for instance, the fact that there is a family level of analysis between household and persons. All in all, the SIPP complex files have up to eight different types of records at five different levels within a single sample unit, and a number of pointers are required to describe all of the relationships. This structure is not particularly easy for a nonprogrammer to use, particularly for someone who wants to use statistical software packages. Processing information on the SIPP complex file is discussed in the next section.

Because some users may find the complex structure difficult to use, the Census Bureau also makes SIPP microdata available in rectangular form. A rectangular file has one type of record of consistent length and fixed format throughout. (The term "rectangular" relates to the fact that the data fields line up column by column and record by record, suggesting a rectangle--see figure 4.3.) Thus, the same data item is found at the same location on each record in the file.

FIGURE 4.3 - COMPONENTS OF RECTANGULAR RECORD
(Example: Wave 4--Rotation Group 1--Interview in October)



The SIPP rectangular file has one record for each person (adult and child) in the sample. All of the subject matter on the complex file is also presented on the rectangular file.[1] The various record types on the complex file constitute segments of the person record on the rectangular file. Sample unit information appears on the record of each person in the unit. Household, family, and subfamily characteristics also appear on each person's record with further differentiation month by month. Each income type, which may be associated with a unique income record on the complex file, is given a unique set of data fields on the rectangular record.

[1] A few items without any particular analytic utility, e.g., interview "book" number, appear on the complex files but do not appear on rectangular files.

The main advantage of the rectangular file is its ease of use. There is no nested structure to deal with, nor does the user need to manipulate the system of pointers to relate the various levels of analysis (households, families, persons, etc.) to each other. Theoretically, the complex file could be more economical in terms of space than a rectangular file, since space does not need to be provided for items for which there are no data. Unfortunately, different kinds of computers use different conventions for representing variable length records, and only with fixed-length records can the Bureau create a product which can be used by all. Several widely used software packages are much easier to use with fixed-length records as well. Therefore, each record on the complex file is padded out to a fixed length which substantially increases the size of the files.

The following sections describe the two file structures in more detail. Other aspects of working with the files are discussed in the next chapter, which is about use of SIPP for estimation.

STRUCTURE OF THE SIPP COMPLEX FILE

In discussing this version of the cross-sectional files, a review of each of these record types--sample unit, household, family, person, and income--and the type of data found on each is provided. (Figure 4.4 summarizes the type of information included on each type of record.) Then the actual sequence of records is discussed, with notation as to how this sequence differs from a hierarchical sequence.

Sample Unit Record

There is one sample unit record for each ADDRESS selected into the SIPP sample at Wave 1. There are 19,878 sample unit records on the 1984 Panel.[2] To each sample unit is assigned a part of the identification number described in chapter 5. The sample unit ID is a unique identifier that becomes part of the address and person identification numbers. Also at this level is the identification of the state in which the original sample address was located (with some states grouped for confidentiality reasons) and a "reduction group code" for use in subsampling.

The sample unit record contains counts of all other records associated with the sample unit during the wave by record type, i.e., the number of household, family, person, and income records. All of these records follow the sample unit record.

Address Record [3]

There is one record per address occupied by at least one original sample member during the reference period of the wave. Since SIPP follows original sample

---

[2]Starting with Wave 5 of the 1984 Panel, the number of sample units is reduced by about 3000 as described in chapter 5.

[3]Most references to this record type in the SIPP technical documentation use the term "household record." This is misleading if the user is interested in the household as an analytic unit because the record really pertains to a physical address.

FIGURE 4.4 - SUMMARY OF RECORDS ON THE SIPP COMPLEX FILE

Record Type	Type of Information Included
Sample unit 1 record per sampling unit	Numbers of households, families, and persons associated with the sample unit State code of original sample unit
Household 1 record per address at which sampled persons lived during 4 months	Numbers of associated families and persons Half-sample codes for variance estimation Geographic codes: state, metropolitan subsample Household weights by month (4 reference months and interview month) Type of unit, public housing and rent assistance Aggregate household income by month by type Participation in school lunch and energy assistance programs
Family Month 1 record per family per month	Type of family (defined in discussion of family record) Number of persons in family Aggregate family income by type [1] Family Weight
Person 1 record per person	Basic demographic characteristics Month in sample codes Month-to-month change flags for household, family, and subfamily Aggregate income by month by type Labor force status by week and month Identification of programs from which income is received
Wage/Salary Records for up to 2 jobs per person	Occupation and industry Weeks worked and hours worked Reported and imputed wage/salary income by month
Self-employment Records for up to 2 businesses per person	Occupation and industry Weeks worked and hours worked Form of ownership, number of employees Reported and imputed income from business by month Net profit/loss
General Amounts Type 1 Records for up to 6 sources per person	Type of income (see figure 3.1) Reported and imputed income from this source by month Pointers to other persons covered by payment (if applicable)
General Amounts Type 2 0 or 1 record per person	Amount of reported income from 13 types of assets (individual and joint--see figure 3.1) Constructed monthly imputed amounts for each individual

[1]The summary income fields are total cash income, earnings, asset income, means-tested cash transfers, other unearned income, Social Security, SSI, unemployment compensation, veterans' compensation and pensions, AFDC, or food stamps.

persons when they move, a new address record is added to the file when one or more original sample persons leave the original sampled address and move to a new address. For example, when the daughter and her husband in figure 2.4 moved out of the parental home in Wave 4, the resulting Wave 4 file contained two address records, one for the parental home and one for the new home of the daughter. If all members of a household in one wave relocated to a new address during the 4-month reference period of a subsequent wave, two household records would appear on the file in the wave in which the move occurred. One address record would contain characteristics of the old address and denote the months to which they are applicable, and the second address record would carry characteristics of the new address. As an example, when the parents in figure 2.4 moved to a new address in Wave 7, two address records were generated for them. One pertained to their old address and one pertained to the new. In the absence of any other movement, files for later waves would only have one address record pertaining to the new address. In the Wave 1 public-use files there is only one address comprising the sample unit because household composition is fixed as of the interview date, regardless of what changes may have taken place in the 4 prior reference months. However, in later waves, if original sample members move and establish new households at new addresses, additional address records are created as illustrated above.

Each address record is uniquely identified by a combination of the nine-digit sample unit ID and a two-digit current address ID, which differentiates addresses within the sample units. (The SIPP ID system is explained in chapter 5.) For example, when the daughter and son leave in Wave 4 in figure 2.4 and move to a new address, the sample unit ID would be carried from the household that she maintained with her parents in Wave 1, but a new current address ID is assigned in Wave 4 for the address she shares with her husband.

Some information about the household is also on the address record: 1) items pertaining to the interview date, i.e., items for which monthly values are not collected, such as tenure, rent, and energy assistance; 2) household weight by month; and 3) derived summary characteristics about household members collectively--most particularly--total household income and its components for each month of the reference period. The address record indicates whether, in each month, the household's income included earnings, means-tested cash or noncash transfers (such as AFDC or food stamps), nonmeans-tested benefits (such as Social Security), or asset income. (Monthly household amounts for each of these income types also appear on this record.)

Family Month Record

The family as defined by the Census Bureau is two or more persons related by birth, marriage, or adoption and living together. In SIPP, family units are defined each month and a separate record is created for each family in each month. If the concept of families was interpreted strictly, persons living alone or with nonrelatives, for instance, would not be living in families. As the family level of analysis is used to cover everyone in the SIPP files, however, records are constructed for individuals not living with a relative. These one-person families are called pseudo-families. There can be as many as eight pseudo-families living together; nine constitutes group quarters. Thus, a family which has a foreign student living with them would be treated as two family units in SIPP files, one family consisting of the related persons and a second pseudo-family consisting of the unrelated student.

For Wave 1, family composition is fixed as of the interview date; changes during the reference months are not measured even though separate family month records exist in the Wave 1 files. For Wave 2 and beyond, family composition may change from month to month--families may dissolve (a husband dies) or be created (two individuals marry) from one month to the next--and these changes are recorded by SIPP. This richness in detail leads to some complexities in defining what constitutes a single family over time. The structure of this file sidesteps these issues by defining families independently each month. Thus, the family month record on the complex files refers to only one reference month.

The family month record contains information on type of family. Family types and definitions are as follows:

Primary family	The household reference person together with all other household members related to him/her.
Subfamily	A married couple and dependents related to the reference person but not including him/her, such as the householder's daughter and son-in-law, or a single parent-child group where the child is under 18 and the group is related to the householder.
Unrelated Subfamily	Family group not consisting of a relative of the householder, but living in the household, such as a boarder and his/her spouse. This is sometimes referred to as a secondary family.
Primary individual	Reference person who does not have any relatives living with him/her. (This is a pseudo-family.)
Secondary individual	Household member unrelated to the reference person or to anyone else in the household. (This is a pseudo-family.)

SIPP contains one family month record for each primary family, unrelated subfamily, primary individual, and secondary individual residing at the address each month. (NOTE: Primary family month and primary individual family month records will not occur simultaneously within a household in one month.)[4] When subfamilies are present, there will be separate subfamily month records as well as a primary family month record. Persons in related subfamilies are represented in both the primary family month and subfamily month records. For example, total family income and numbers of persons in the family include related subfamily members and their income in both the primary family month and subfamily month records.

---

[4]In any given month there is at least one household reference person and, hence, at most, one family containing the household reference person. If the household reference person has other relatives present, then this family record represents a primary family. If not, this (pseudo) family represents a primary individual.

Kind of family (husband/wife, no spouse present, or individual) and number of children are also on the family month record. To obtain more detail on family relationships and composition, one would need to examine each person record that is associated with the family for the month. For example, to find three-generation families, it would be necessary to compile relationship information from the person record for each family member.

Each family record contains aggregate total monthly income for the following types: earnings, property income, government transfer cash income, Social Security, SSI, unemployment compensation, veterans benefits, AFDC, and food stamps. (Chapter 3, "Survey Content," describes income sources more fully.)

Person Record

One record appears on the file for each original sample person and for each additional person who resides with original sample members during the reference period of the wave. The person record is organized around the SIPP questionnaire, except for amounts of income received and constructed monthly demographic and program participation variables. The person record is much more detailed than the sample unit, household, or family month record.

Several basic demographic characteristics are on the person record. Unedited relationship to the reference person, person number (see chapter 5) of spouse and parents, and marital status are recorded from the previous wave (except Wave 1). Edited and unedited data are constructed for each month of the (present wave) reference period and recorded as of the interview date. Birth year is given, as is age for each reference-period month and as of the interview date. The person's race, sex, and detailed ethnic origin are presented, as well as highest educational level completed. For new survey entrants, the reason they entered is given (birth, marriage, other, or at the address prior to the original sample person); for previously interviewed persons no longer at a sample address, the reason for their leaving is also indicated (deceased, institutionalized, living in Armed Forces barracks, moved abroad, separated or divorced, no longer with an original sample person, or other). Date of entering or leaving is also shown. Note that persons who move from one address to another will have both "enter" and "left" dates recorded.

In order to facilitate identification of when persons enter and leave the sample, month-in-sample codes are constructed for each of the 4 reference months and the interview month. In a given month a person is identified as "in sample" or "not in sample" depending on whether she or he resided at an interviewed address for more than one-half of that month. For example, the mother, father, and daughter in figure 2.4 will have month-in-sample codes set to "in sample" for every month. The son who joined the Army in Wave 2 will be classified as "in sample" in Wave 1 and the early months of Wave 2, but classified as "not in sample" in the rest of Wave 2 after he left. The son will have no record in subsequent waves and, hence, will not be classified at all. The husband who appears in Wave 3 will be classified as "not in sample" in the months before the wedding and "in sample" after the wedding.

These month-in-sample codes are important to consider when using demographic data in the person record. In the months when a person is classified as "not in sample," the monthly demographic data will be "0." Valid codes will appear in the other months.

The person record indicates whether the adult received income from any of the different sources described in figure 3.1. Total income and income in five broad categories (earnings, asset income, means-tested cash transfers, and total other income) are constructed for each reference month. There is a consolidated set of recipiency and asset ownership indicators (INCSOURC and ASTSOURC), but amounts from each detailed source are recorded in the income records that follow the person record on the file.

Labor force activity is another subject on the person record. Employment status (with or without a job, on layoff, looking for work) and weeks with a job and/or without pay are shown for each month and week of the reference period for each adult. Whether an adult respondent without a job was not looking for a job or why he or she could not take a job in those weeks is included, as is part-time work (less than 35 hours per week).

Program coverage fields are constructed and placed on the person record. These files denote coverage under Medicare and Medicaid; health insurance; veterans payments; Women, Infants and Children Nutrition Program (WIC); AFDC; food stamps; general assistance; foster child care payments; Social Security; and Railroad Retirement. Coverage is specifically recorded for each reference month for both adults and children.

With the exception of WIC, benefits received by adults under nonhealth care programs are recorded in the income records. WIC benefits are recorded on the person record; unlike other income and benefit sources, amounts are recorded for both adults and children. Also, unlike the other income sources, the value of the WIC benefit was imputed (these were not collected in the interviews) and the adult's benefits does not include the child's amount.

Receipt of government or private pension income or disability payments by adults is indicated on the person record, as is receipt of alimony or child support. Amounts of benefits received, however, appear on income records.

Ownership of various types of assets and savings accounts or investments are shown near the end of the adult's person record, while corresponding amounts appear on an income record. A series of items on educational assistance complete the person record.

Income Records

With only a few exceptions, there is a separate income record for each source of income for each person. Depending on the type of income, the information may be recorded in one of four formats or types of records:

- Wage and salary income (up to two records per person)
- Self-employment income (up to two records per person)
- General Amounts Type 1: Nonearnings income from programs and miscellaneous sources (up to six records per person)
- General Amounts Type 2: Nonearnings income from owned assets (up to one record per person with asset income)

Thus, if a person had two wage or salary jobs during the 4 months, either simultaneously or one after the other, there would be two wage/salary income records. If a person had income from three different program or miscellaneous income sources (for example, unemployment compensation, AFDC, and child support payments), there would be three General Amounts Type 1 records. The General Amounts Type 2 record consolidates asset income from a number of sources, including assets like savings accounts or rental property.

Income questions are asked only of persons 15 years old and older, so there are no income records for children. Some types of income from which children benefit appear on a corresponding adult's record, e.g., food stamp income. SIPP misses a small amount of income received solely by children, e.g., from a job delivering papers or interest from savings.

Income from more than two wage or salary jobs or more than two self-employed jobs is recorded on a General Amounts Type 1 record coded as "other cash income not included elsewhere" (ISS code = 56). If the respondent receives income from six programs, the miscellaneous, catch-all income source, "other cash income not included elsewhere," will be missed. Such a large number of income sources for any individual is expected to be extremely rare, however.

1. Wage and Salary Record

The wage/salary (WS) record is obtained from Section 2 (Earnings and Employment), part A of the SIPP questionnaire. Each person 15 years old or more who worked for an employer during the reference period has an associated WS record. Along with information on the respondent's occupation and industry, the WS record shows how many weeks the person was employed and how much he or she earned each reference month.[5] The WS record also indicates type of employer (private, government, Armed Forces, unpaid in family business or farm). Hours worked per week and hourly rate of pay are also given.

If the person worked for more than two employers, records are created for the two jobs at which he or she worked the most hours. Employment at odd jobs or for several employers (such as domestic work) is recorded as a single job type on one WS record. As noted above, income from more than two jobs is recorded on a separate General Amounts Type 1 record as "other cash income not included elsewhere" (ISS code = 56).

2. Self-Employment Record

Deriving from part B of section 2 on the SIPP questionnaire, all self-employed adult respondents are given self-employment (SE) records. Occupation and industry are shown. Also shown for each reference month are the number of weeks with the business; whether the person received income from the business; and, if so, how much. The SE record displays hours worked at the business per month.

Employment size of the business and form of organization (corporation, sole proprietorship, partnership) are on the SE record. It also shows whether other

---

[5]Due to the manner in which they were created, there are inconsistencies between weeks worked by month per job for the wage and salary record and weeks worked by month from the person record.

family members were owners or partners in the business and how much income they received from it. Self-employment draw is recorded monthly in each self-employment record. Other information on the SE record includes an estimate of profit or loss during the reference period and net earnings.

As is true for wages and salary jobs, income from more than two self-employed jobs is recorded on a separate General Amounts Type 1 record as "other cash income not included elsewhere" (ISS code = 56).

3. General Amounts Type 1 Record

This record covers nonearnings income from transfer programs and certain other sources, referred to below. Each person receiving these types of incomes can have up to six General Amounts Type 1 (G1) records on the file, one for each type of income source. If more than six of these income sources apply to a respondent, the questionnaire is filled out for the first six in numerical order from the Income Source Summary, shown in figure 3.1 of the Users' Guide. Income from remaining sources is not recorded. The complete list of sources for which recipiency was indicated can be found on the person record (refer to INCSOURC and ASTSOURC).

Each G1 record specifies the recipient, the type of income (ISS codes 1 to 75 from the Income Source Summary), and the amount received in each month of the reference period. The G1 record also shows receipt of benefits by or for a child in the household for social security and railroad retirement income. Relevant G1 records specify the individuals (by person number) covered by veterans' payments, AFDC, general assistance, foster child care payments, or other welfare; by social security or railroad retirement benefits; and by food stamps. Jointly received railroad retirement and social security benefits have been allocated between spouses with the husband's share equal to 66 percent of the total.

4. General Amounts Type 2 Record

General Amounts Type 2 (G2) includes income amounts from owned assets, including property. A single G2 record for each adult summarizes all sources of income from assets. Income from jointly owned assets is recorded separately from income from those held individually. Income received jointly by spouses is recorded initially on one spouse's record, then prorated in equal shares to each spouse. While some of the types of asset income are recorded for each month, others show only the amount of income received during the entire 4-month period, so that an amount for each month is prorated by the Census Bureau.[6] If the respondent does not know the income amount, he or she is asked the amount invested or held in the asset; the income earned is estimated by the Bureau. All this information except the asset value is included on the G2 record.

[6]The lump sum amount is divided by "4." If it is not evenly divisible by "4," then the remainder is allocated in whole dollar amounts to the early months of the reference period until the remainder is exhausted.

Sequence of Records on the Complex File

For any one point in time, these eight types of records could be described as five levels: sample unit, addresses within sample units, family months within addresses, persons within addresses, and income sources of four types for each adult. However, since family and household membership can change month to month, the structure is not strictly hierarchical and, hence, not ordered in the implied manner. Instead, all records for a sample unit are kept together; but within sample unit, records are sorted by record type. The sample unit record is first, followed by one or more address records, then four or more family records, a person record for each address member, and a variable number of income records ordered by type. All relationships among records are defined by pointers on each record, not by sequence of records.

The overall sequence is illustrated in figure 4.5. Sample units are sorted by interview month (in effect keeping each rotation group together), by state of sampling unit (not necessarily the same as the current state of residence), within rotation group, and by sample unit ID within state (after the sample unit ID has been scrambled to preserve the confidentiality of the respondents).

Example Illustrating Structure of the Complex File

Consider the household in Wave 3 in figure 2.4 consisting of a husband-wife family that included a subfamily composed of their daughter and her husband living with them. Between Waves 3 and 4 the subfamily moved to their own house.

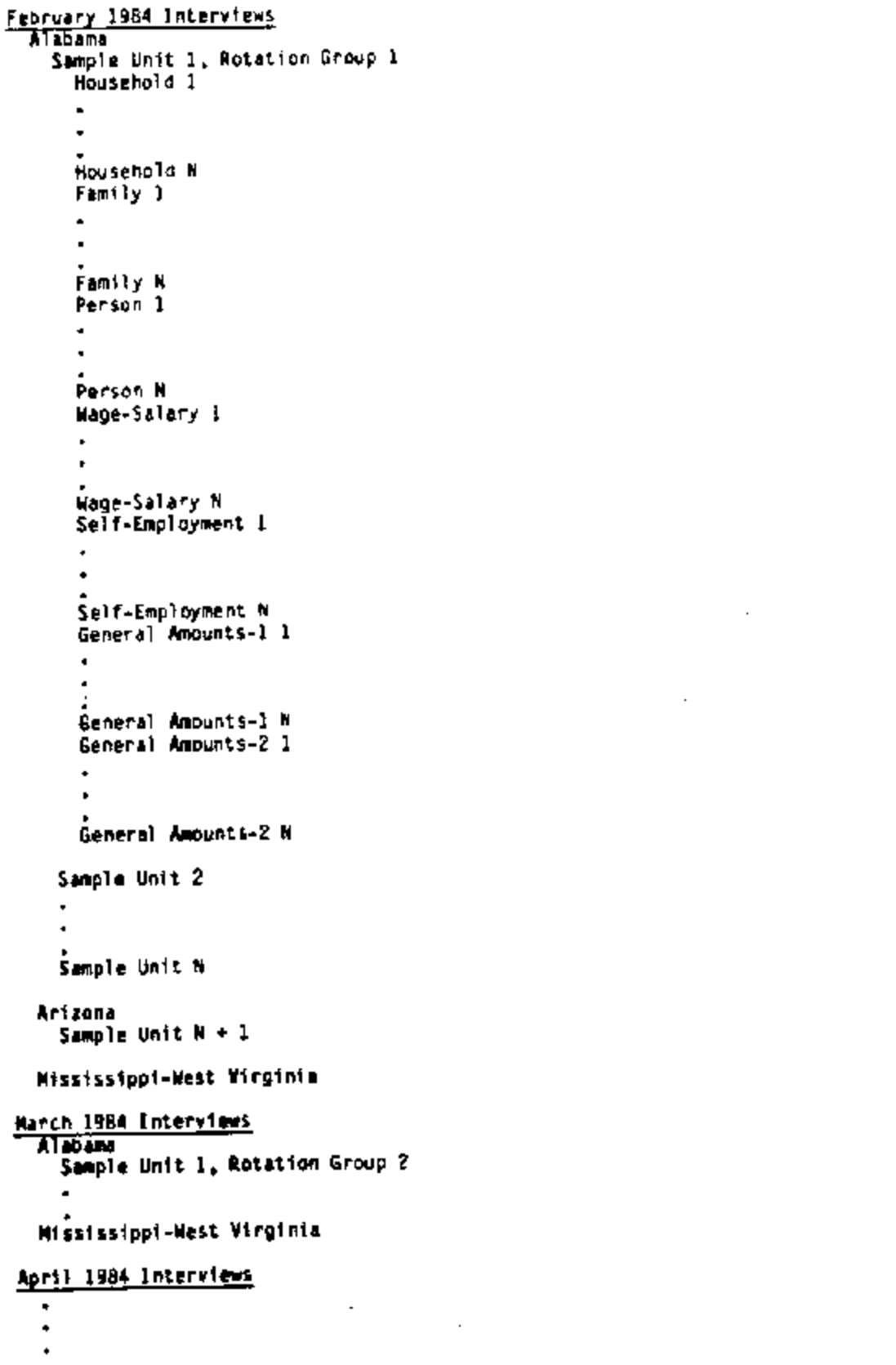
Figure 4.6 illustrates the record sequence for this sample unit in Wave 4, assuming the split occurred in the beginning of August and the interview took place in October. Household 1 (H1) included the father-mother family and the daughter-husband subfamily for the months of June and July. Thus, there is one primary family record and one subfamily record relating to H1 for each of those months. Household 2 (H2) was formed in August when the subfamily left to establish a separate household. Therefore, the first family record for August (F5) relates to household 1 and the second family record for August (F6) relates to Household 2. Pointers on the household record tell which family and person records belong to that household for the specified month. Pointers on the family record tell which person records belong to that family for the specified month, and pointers on the person record indicate which household(s) or family(s) that person belongs to each month. (Subfamily members belong to both the subfamily and the primary family, while other individuals only belong to one family.)

Let's look at the example in greater detail. The above example illustrates record structure in Wave 4 reflecting an interview in October. When the household was interviewed in Wave 1 in October of the previous year, it was found to be made up of five persons. They were assigned entry address ID's and person numbers as follows (see chapter 5 for more information concerning the identifiers):

Entry Address/Person Number Relationship

11	101	Reference person
11	102	Spouse
11	103	Daughter
11	104	Son
11	105	Cousin

FIGURE 4.5 - SEQUENCE OF RECORDS ON THE COMPLEX FILE: 1984 PANEL, WAVE 2 EXAMPLE



Between Waves 1 and 3 person 104 (the son) had left and the daughter's husband moved in. The husband was assigned an entry address ID of 11 (because that was the address to which he entered) and assigned a person number of 301 because he entered in Wave 3).

FIGURE 4.5 - RECORD SEQUENCE ON THE COMPLEX FILE: 1984 PANEL, WAVE 4 EXAMPLE

	Wave Pointers	Monthly Pointers Jun.	Jul.	Aug.	Sept.
S1 Sample Unit					
H1 Household		F1	F3	F5	F7
		F2	F4	P1	P1
		P1	P1	P2	P2
		P2	P2	P5	P5
		P3	P3		
		P4	P4		
		P5	P5		
H2 Household				F6	F7
				P3	P3
				P4	P4
F1 (Jun) Family (of H1)					
F2 (Jun) Family (of H1): Subfamily Record					
F3 (Jul) Family (of H1)					
F4 (Jul) Family (of H1): Subfamily Record					
F5 (Aug) Family (of H1)					
F6 (Aug) Family (of H2)					
F7 (Sep) Family (of H1)					
F8 (Sep) Family (of H2)					
P1 Person 101 - Reference Person	I5	H1	H1	H1	H1
		F1	F3	F5	F7
P2 Person 102 - Spouse	I1	H1	H1	H1	H1
		F1	F3	F5	F7
P3 Person 103 - Daughter	I2	H1	H1	H2	H2
	I3	F1	F3	F6	F7
		F2	F4		
P4 Person 301 - Daughter's Husband	I4	H1	H1	H2	H2
	I6	F1	F3	F6	F7
		F2	F4		
P5 Person 105 - Cousin		H1	H1	H1	H1
		F1	F3	F5	F7
I1 Income: Wage and Salary (of P2)	P2				
I2 Income: Wage and Salary (of P3)	P3				
I3 Income: Wage and Salary (of P3)	P3				
I4 Income: Self-Employment (of P4)	P4				
I5 Income: General Amounts Type 1 (of P1)	P1				
I6 Income: General Amounts Type 2 (of P4)	P4				

In Wave 4, in October, the household, located in New York, had only persons 101, 102, and 105 present, persons 103 and 301 having moved to California in August. The appropriate information was dispatched to an interviewer in California, who located the couple and obtained information about their new household. This information was placed with the same sample unit in the data file during processing at the Census Bureau. Even though persons 103 and 301 (daughter and husband) formed a separate household during the Wave 4 reference period, they retain the same sample unit identifiers, entry address identifiers, and person numbers that they were originally assigned.

Income information for the household was found to be as follows:

Person 101 is retired and receives Social Security

Persons 102 and 103 are salaried

Person 301 is self-employed, and also earned interest on a savings account

Person 105 is under 15 and, therefore, has no income records.

There is a separate income record for each source, as illustrated in figure 4.6. Income of the father (P101) from Social Security is found on a General Amounts Type 1 record. The mother (P102) and daughter (P103) have salaried jobs; their incomes are on separate wage/salary records. In fact, since the daughter had one job in New York, then took another in California, she has two wage/salary records. The husband (P301) is in business for himself, and has a self-employment record. He also has income-producing assets, found on a General Amounts Type 2 record. Note also that the income records are not sequenced by household or family, only by type of record. There are pointers to the first and last income record by type on each person record. Similarly, on each income record, there is a pointer to the person who received the income. Note that these pointers do not vary by month because there is at most one income record per income source per person in any given wave.

It is illustrative to note the impact of the timing of the household split in the preceding example relative to the timing of the SIPP interviews. Because the daughter resided with the parents at the time of the Wave 1 interview, their relationships are known and all three are tracked over time. However, the situation would have appeared differently had the split occurred before Wave 1. In that case, only information pertaining directly to the people in the parental home would have been ascertained. In fact, the user would not have any information on the daughter, including her circumstances and her relationship to the parents, unless her address was selected into the sample along with her parents. Even if her address was selected, her relationship to the parents would not have been measured.

STRUCTURE OF THE RECTANGULAR FILE

In the rectangular version of the SIPP core files, the person is the unit of analysis. There is one very long record for each person, about 5,400 characters in all, containing all of the applicable information about the person: sample unit, address(es) and monthly household, family, personal characteristics, and

income sources. The rectangular file carries all the substantive information found on the complex file. An illustration of the rectangular file is provided in figure 4.8 which appears later in this chapter.

The rectangular structure results in some redundancy. For instance, information which is the same for each member of an address is repeated on each person's record. Extra space is also required to provide a unique location for each possible source of income, when most individuals receive only a few types of income. The result is a structure much easier to program for or to use with common statistical software programs than is the complex file. The rectangular format is also more amenable to resorting and restructuring, as in the analysis of food stamp recipiency units, or the grouping of persons covered under the same health insurance policy.

The sequence of segments on the rectangular record generally follows the sequence of records on the complex file, as illustrated below.

Sample Unit Segment

Sample unit information appears on the first part of the record and contains basic data about what appears on the other portions of the record as well as the sample unit identification codes. The sample unit segment derived from the sample unit record of the complex file gives the sequence number of the sample unit on the file and the sample unit identifier. (See chapter 5 for an explanation of the derivation of the sample unit identifier.) The rotation group is given (1, 2, 3, or 4), as well as the number of addresses associated with that sample unit in each month of the reference period and, as of the interview date, the number of family month records, and the number of person records for the sample unit. This segment also gives the total number of wage and salary jobs and self-employment businesses for members of the sample unit, and the number of sample unit adults who receive asset-based income.

A state code is given for most sample units (see figure 4.7). (See chapter 5, "Use of Cross-Sectional Files for Estimation and Analysis," for appropriate use of state codes.)[7] There is also a "reduction group code" for use in reducing sample size if that becomes necessary. This code was used by the Census Bureau for the sample cuts implemented in SIPP.

Address Segment

Address information is in five pieces--one for each of the 4 reference months and a short piece for the interview date. The same sequence of data items is repeated for each of the 4 reference months. Beginning with month 1 (the earliest month of the reference period), the address segment gives the address identification number, the calendar month, current address ID, and the entry address ID and person number of the reference person.[8] Base weight and final

---

[7]Some states are combined for confidentiality reasons on the public-use tapes.

[8]Noninstitutional group quarters, such as rooming and boarding houses, college dormitories, and convents or monasteries, consist of a group of unrelated individuals or one or more families. They are interviewed if the group quarters is their usual residence. Therefore, there is no reference person and the first person is listed for the address.

FIGURE 4.7 - GEOGRAPHIC CODES USED IN SIPP

State Codes

01. Alabama
*02. Alaska
04. Arizona
05. Arkansas
06. California
08. Colorado
09. Connecticut
10. Delaware
11. District of Columbia
12. Florida
13. Georgia
15. Hawaii
17. Illinois
18. Indiana
19. Iowa
20. Kansas
21. Kentucky
22. Louisiana
23. Maine
24. Maryland
25. Massachusetts
26. Michigan
27. Minnesota
29. Missouri
*30. Montana
31. Nebraska
*32. Nevada
*33. New Hampshire
34. New Jersey
36. New York
37. North Carolina
*38. North Dakota
39. Ohio
40. Oklahoma
41. Oregon
42. Pennsylvania
44. Rhode Island
45. South Carolina
47. Tennessee
48. Texas
*49. Utah
*50. Vermont
51. Virginia
53. Washington
55. Wisconsin
90. Idaho, New Mexico, South Dakota, Wyoming
91. Mississippi, West Virginia

*These seven states are included as states to which households moved.

Identifiable MSA/CMSA Codes as defined in 1983

0160 Albany-Schenectady-Troy, NY
0640 Austin, TX
0840 Beaumont-Port Arthur, TX
1000 Birmingham, AL
1122 Boston-Lawrence-Salem, MA-NH CMSA (part in MA only)
1282 Buffalo-Niagara Falls, NY CMSA
1602 Chicago-Gary-Lake County (IL), IL-IN-WI CMSA (parts in IL and IN only)
1642 Cincinnati-Hamilton, OH-KY-IN CMSA (parts in OH and KY only)
1692 Cleveland-Akron-Lorain, OH CMSA
1840 Columbus, OH
1922 Dallas-Fort Worth, TX CMSA
2000 Dayton-Springfield, OH
2082 Denver-Boulder, CO CMSA
2162 Detroit-Ann Arbor, MI CMSA
2400 Eugene-Springfield, OR
2840 Fresno, CA
3120 Greensboro--Winston-Salem--High Point, NC

FIGURE 4.7 - CONTINUED

3240 Harrisburg-Lebanon-Carlisle, PA
3282 Hartford-New Britain-Middletown, CT CMSA
3320 Honolulu, HI
3362 Houston-Galveston-Brazoria, TX CMSA
3480 Indianapolis, IN
3762 Kansas City, MO-Kansas City, KS CMSA (part in MO only)
4000 Lancaster, PA
4472 Los Angeles-Anaheim-Riverside, CA CMSA
4720 Madison, WI
4920 Memphis, TN-AR-MS (part in TN only)
4992 Miami-Fort Lauderdale, FL CMSA
5082 Milwaukee-Racine, WI CMSA
5160 Mobile, AL
5360 Nashville, TN
5480 New Haven-Meriden, CT
5560 New Orleans, LA
5602 New York-Northern NJ-Long Island, NY-NJ-CT CMSA
5880 Oklahoma City, OK
5920 Omaha, NE-IA (part in NE only)
5960 Orlando, FL
6152 Philadelphia-Wilmington-Trenton, PA-NJ-DE-MD CMSA (parts in PA, NJ, and DE only)
6200 Phoenix, AZ
6282 Pittsburgh-Beaver Valley, PA CMSA
6442 Portland-Vancouver, OR-WA CMSA (part in OR only)
6840 Rochester, NY
6920 Sacramento, CA
7042 St. Louis-E. St. Louis-Alton, MO-IL CMSA (part in MO only)
7080 Salem, OR
7240 San Antonio, TX
7320 San Diego, CA
7362 San Francisco-Oakland-San Jose, CA CMSA
7560 Scranton--Wilkes-Barre, PA
8000 Springfield, MA
8280 Tampa-St. Petersburg-Clearwater, FL
8400 Toledo, OH
8520 Tucson, AZ
8560 Tulsa, OK
8840 Washington, DC-MD-VA
8960 West Palm Beach-Boca Raton-Delray Beach, FL
9240 Worchester, MA
9280 York, PA
9320 Youngstown-Warren, OH

weight follow (see Chapter 7, "Assessing Reliability of SIPP Data," for a discussion of the difference), as do the half-sample and stratum codes for variance estimation. Next are geographic codes for state of current residence; whether the address is in the metropolitan subsample (see chapter 5); and, if so, the metropolitan area of residence (see figure 4.7).

Various items from the control card follow. These include items that describe the housing unit or group quarters, and whether the unit is in a public housing project or benefits from rent subsidies.

The address segment also contains aggregate income information for each month. Total household income, earned income, property income, means-tested cash transfers, and noncash income (through food stamps, WIC, and energy assistance) are given. Social Security, Supplemental Security Income (SSI), unemployment compensation, veterans payments, AFDC, and food stamps are also shown as monthly household aggregates. Participation of the address unit members in energy assistance programs and subsidized school breakfasts and lunches are recorded as well.

The above information is then repeated for months 2, 3, and 4 of the reference period. A few household characteristics are recorded as of the interview date, as well as for each month of the reference period. These include household size, sample weight, current address ID, and entry address and person number of the reference person.

If the composition of the household at each address changed or if the person moved to a new address, constructed information on each monthly household--household size, total income, etc.--would change to reflect the composition of the household of which he or she was a member that month. Other information which is collected only at the time of the interview--tenure, public housing, school lunch participation, etc.--remains constant unless there is a change in address. (Any change in composition or location would also be highlighted in the person segment of the record by a household change indicator referring to each successive pair of months. This is discussed further in the section on the person segment.) The two-digit current address ID and geographic codes would change when the respondent moves to a new address.

Family Month Segment

The family month segment consists of four parts rather than five as in the address segment, each one corresponding to a reference month. (Interview month data are omitted.) The information, such as family type (see above discussion of family records on the complex file for definition), number of children, and family income by type, is the same for every family member. For unrelated individuals (or pseudo-families), data in the family segment refer to that individual alone. Thus, a tabulation of data for families must exclude data for unrelated individuals which appear in the family data fields. Unrelated individuals are identified as primary or secondary individuals in the family type variable.

There is no family numbering system which allows the user to trace data for the same family across the 4 months on either of the two types of files. The rectangular file (person segment) includes a family composition change indicator similar to that for households, which identifies whether any change occurred

between each successive pair of months either in membership, address, or relationship. (This is discussed further under the person segment.) One cannot tell from that flag whether the change has disrupted the family as a unit, as in the death or divorce of the reference person or spouse, or whether it has not, as in the birth of a child. Note that there are only three operative change flags, occurring between each successive pair of months within the 4-month period. There is no indication of whether the family composition changed between the last reference month of the previous interview and the first reference month of the current interview.

Subfamily Month Segment

While the complex file gives subfamilies a "subfamily month record," the rectangular file assigns subfamily information to a separate set of fields corresponding directly to the four monthly family segments. For persons who are members of subfamilies, the subfamily segments characterize that unit for a given month in terms of type, number of children, and the same income aggregates as on the household segment. These figures are also included in the totals shown on the corresponding primary family segments since subfamily members are also, by definition, members of the primary family. The user must, therefore, be wary of double-counting subfamily income. The person segment contains an indicator of change in subfamily membership from month to month (discussed under the person segment below).

Person Segment

The person segment, which is derived from the person record of the complex file, records the characteristics of each person over the 4-month period. Unlike the address, family month, and subfamily month segments, the data are not grouped by month; rather characteristics for different months appear in successive data fields for each variable. The organization of the person segment is similar to the control card and questionnaires, except that constructed monthly demographic program participation and amounts of income received are added.

For a few characteristics--address ID, relationship to reference person, marital status, and person numbers of spouse and parent (if present)--unedited data are carried over from the previous wave (except for Wave 1). This information, recorded as of the date of the previous interview, should generally be consistent with corresponding characteristics for the first reference month of the current interview. Still, information for the first reference month refers to the characteristic as it was for most of the month; if the previous interview occurred early in that month, there could be some discrepancies; e.g., if the person was single at an interview on the 6th of the month, but married on the 12th of the month. In any case, the data which appear for the previous wave help the user assess consistency of the data.

Like the person record on the complex file discussed earlier, the person segment contains month-in-sample indicators, monthly demographic data (which are zero for months not in sample), reason and dates entering and leaving the

sample,[9] labor force activity, program coverage, income recipiency, asset ownership, and educational assistance. As noted earlier, WIC benefits are imputed rather than collected and they are included in the person segment along with the program coverage fields. Unlike other benefits recorded in the income segment, WIC benefits are recorded for both children and adults covered under the program.

As mentioned above, the rectangular file carries a series of household, family, and subfamily change indicators. For Wave 2 of the 1984 Panel, for example, the variable HCHANGE1 is a one-character field located at character 2220. This is followed by HCHANGE2, HCHANGE3, HCHANGE4, and HCHANGE5, for months 2, 3, 4 and the interview month. The value of HCHANGE1, representing household composition in month 1 of the reference period, will always be 1. If no change in household composition or address occurred during month 2, HCHANGE2 will also have a value of 1; if the household changed in composition or address, its value will be 2. Each time a change takes place, the value of HCHANGE will be one greater than in the preceding month. Similar variables follow directly for family composition (FCHANGE1-4) and subfamily composition (SCHANGE1-4), with the difference that these variables do not include the interview month.

The person segment includes a large number of fields which reflect screening for the receipt of various types of income. In Wave 1, respondents 15 years and older are asked about receipt of the various types of program income discussed in chapter 3. In Wave 2 and subsequent interviews, the interviewer checks that income sources (ISS codes 1 to 56 on the Income Source Summary) were correctly recorded in the previous interview, whether respondents still obtained income from each of those sources during the 4 reference months for this interview, and whether any other sources were added to those previously recorded. The same type of checklist is asked about income-earning assets (ISS codes 100 to 150).

The person segment shows aggregate income from all sources and separate aggregates for earned income, property (asset) income, transfer income, and other sources. These aggregates are given for each reference month. There is a consolidated set of recipiency and asset ownership indicators, called INCSOURC (in Wave 1 of the 1984 Panel, characters 2701 to 2720; in subsequent waves of the 1984 Panel, 2577 to 2596) and ASTSOURC (in Wave 1, characters 2721 to 2759; in subsequent waves, 2597 to 2635).

Income Segments

The income segments on the rectangular file correspond to the information recorded on four different types of income records on the complex file. While the complex file has zero, one, or two wage and salary records, representing up to two different jobs, the rectangular file always carries two successive segments for two wage or salary jobs. Occupation and industry codes are

---

[9] Fields labelled "reason" and "dates entering the sample" describe both entering and leaving as do fields labelled "reason" and "dates left." Entrants and exits can be distinguished, however, because the reason codes are unique for each type of movement.

present for each job, as well as number of weeks worked each reference month.[10] The segment shows receipt and amount of earnings each month. If the person began or left the job during the reference period, the date is shown. Other wage/salary information shown include hours worked per week, pay rate, and amount of pay before deductions. There are flags showing whether any of the amount information has been imputed.

The self-employment segment, likewise, contains data fields for two businesses on all records. Aside from receipt and amount of income each month, the segment has the occupation and industry codes, the number of weeks each month the person was with the business,[11] the legal structure, and whether other household members were owners or partners.

For income from program and miscellaneous sources (General Amounts Type 1), there are 35 separate sets of fields, one for each type of income (even though the questionnaire allows no more than six of these income types to be recorded for any one person). Each set of fields contains monthly recipiency indicators, amounts, and imputation flags. For selected income types, additional information was collected and is recorded along with the basic recipiency and amount variables. The income segment shows income from sources numbered 1 to 75 on the Income Source Summary. Income from owned assets (General Amounts Type 2) is also shown on each record. As noted in the discussion of the complex file, income amounts earned jointly by a husband and wife are initially recorded on only one of the spouse's questionnaire, but are allocated 50-50 in the income segment. Similarly, asset income amounts collected only as 4-month totals are prorated to the applicable months, and jointly received Social Security and Railroad Retirement have been allocated (unevenly) to each spouse.

Users of the rectangular file should pay close attention to the existence of not-in-universe (NIU) codes. When no income record for a particular job or income source existed in the complex file, the associated positions in the rectangular file have been padded. On Wave 1 of the 1984 Panel, zeroes were used to fill the amount fields in this case. However, for subsequent waves, a negative number was used for this purpose. When summing income amounts across records, the user should exercise caution in the treatment of these negative numbers in their calculation.[12]

Sort Sequence of the Rectangular File

As with the complex file discussed earlier, all records for a sample unit are kept together, even if component households are currently in different states.

---

[10] As noted earlier, fields denoting weeks worked per month per job are not necessarily consistent with fields denoting weeks worked per month in the person segment.

[11] Ibid.

[12] Particular caution should be exercised in using AFDC benefits (120AMT*). On Wave 2, 1984 Panel, valid dollar amounts occupy the first five characters of the field (as documented) but the NIU code (-00009) occupies six digits. On subsequent waves, both the valid dollar amounts and NIU code occupy all six digits.

The actual sort is as follows: by interview month (i.e., rotation group), by state of sample unit, by sample unit ID (after scrambling to preserve confidentiality), by entry address ID, by person number. The user could thus obtain a representative subsample by selecting one rotation group, all of whose records lie consecutively on the file, or by using the reduction group code found on the sample unit segment of the record.

Example Illustrating Structuring of the Rectangular File

To illustrate how the rectangular file is organized, consider the hypothetical example in figure 2.4. A listing of the records for this example is found in figure 4.8. In Wave 1, we find a sample unit with a household consisting of father, mother, daughter, son, and cousin (F-M-D-S-C). Each person's record identifies the same sample unit (101111103) and address by month (11 each month). Each record further specifies that he or she is at an address with one household and one family each month, and that each person was in the sample for all 4 months.

In Wave 2, the son left to join the military. His record now shows no address or family in months 3 and 4. Since he was no longer in the sample, his record disappears after this wave.

In Wave 3, the father-mother-daughter-cousin unit was unchanged during months 1 and 2, but the daughter married and her new husband (H) came to live with her and her parents during months 3 and 4. The records of the father, mother, daughter, and cousin show the same address and family during months 1 and 2 because no change had occurred. The new husband's record contains out-of-universe or not applicable codes for these months. For months 3 and 4, the record for each person indicates that there was one household with two families (a family and a subfamily). All five people are flagged as "in sample" for these 2 months. Note the sort sequence is not based on family relationships and the daughter and husband do not appear together. Also note the husband was assigned the sample unit and address of the daughter's household to which he moved.

In Wave 4, the situation remains the same for months 1 and 2. In month 3, the daughter and her husband move to a separate address. Each person's record keeps the same sample unit identifier. All records show there was one household in months 1 and 2, and two households in months 3 and 4. The records also all indicate there were five persons connected with the sample unit in all 4 months. The records of the daughter and her husband have the same address identifiers as do the parents in months 1 and 2, and the new address identifier for months 3 and 4.

Figure 4.8 skips to Wave 7 and shows that four people remain in the sample unit. In Wave 5 the cousin moved out and was not followed because he was under age 15. The four people have the same address in months 1 and 2 that they had in the latter part of Wave 4. In month 3, however, the parents moved and were followed. Their new address (71) is, therefore, assigned in the latter 2 months. All persons retained the original family relationships within the two households each month and were in sample for the full reference period.

FIGURE 4.8 - RECTANGULAR FILE RECORDS FOR THE HYPOTHETICAL EXAMPLE

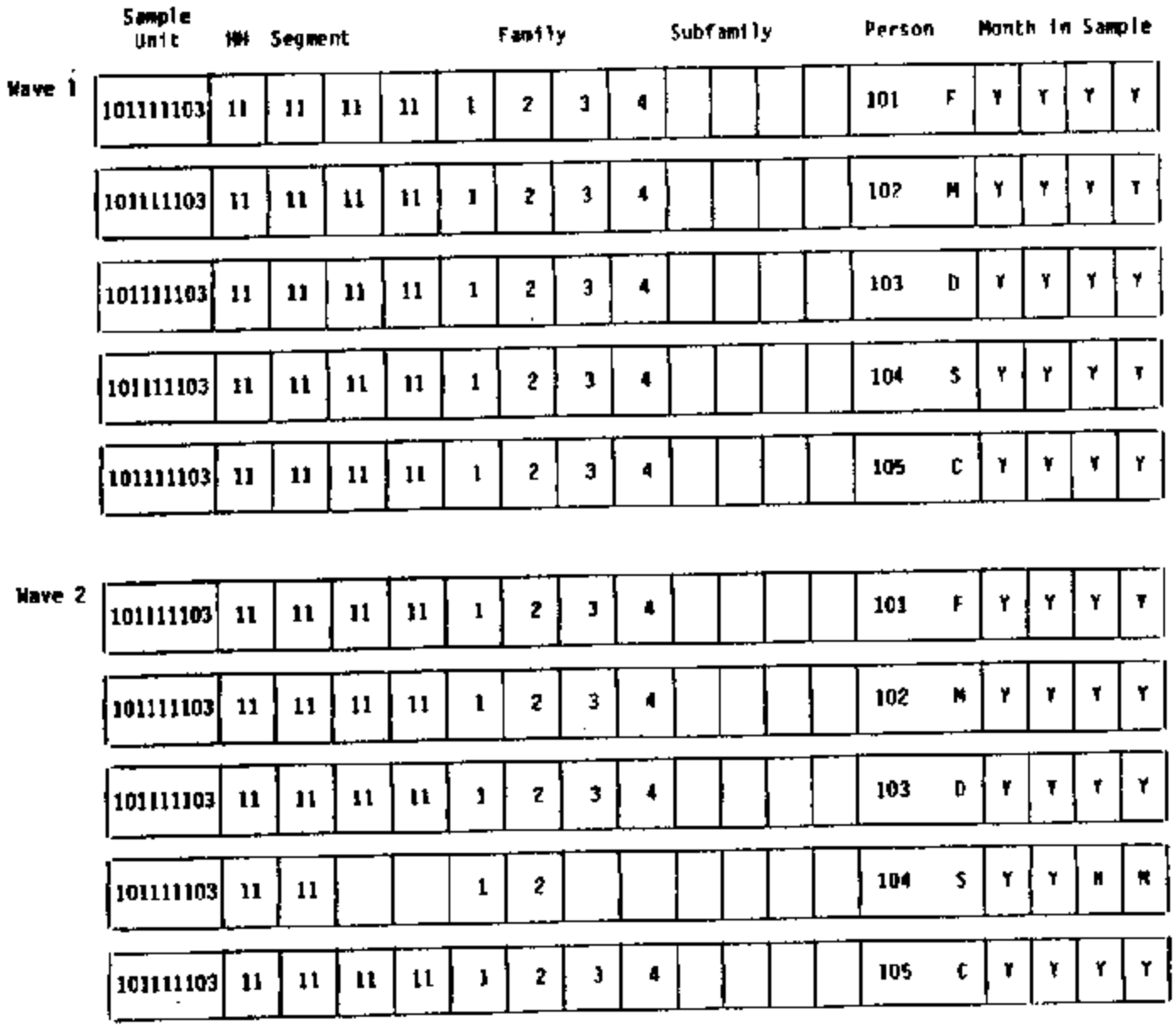
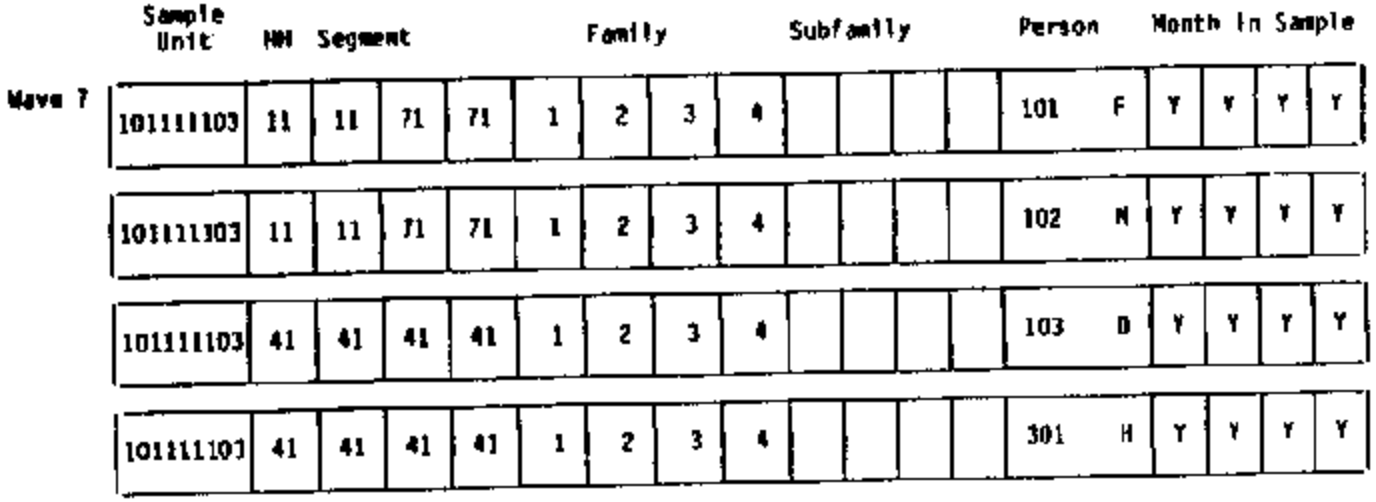


FIGURE 4.B - CONTINUED

	Sample Unit	HH	Segment			Family				Subfamily				Person		Month in Sample			
Wave 3	101111103	11	11	11	11	1	2	3	5					101	F	Y	Y	Y	Y
	101111103	11	11	11	11	1	2	3	5					102	M	Y	Y	Y	Y
	101111103	11	11	11	11	1	2	3	5			4	6	103	D	Y	Y	Y	Y
	101111103	11	11	11	11	1	2	3	5					105	C	Y	Y	Y	Y
	101111103			11	11			3	5			4	6	301	H	N	N	Y	Y
Wave 4	101111103	11	11	11	11	1	3	5	6					101	F	Y	Y	Y	Y
	101111103	11	11	11	11	1	3	5	6					102	M	Y	Y	Y	Y
	101111103	11	11	41	41	1	3	5	6	2	4			103	D	Y	Y	Y	Y
	101111103	11	11	11	11	1	3	5	6					105	C	Y	Y	Y	Y
	101111103	11	11	41	41	1	3	5	6	2	4			301	H	Y	Y	Y	Y

FIGURE 4.8 - CONTINUED



READING THE DATA DICTIONARY

Figure 4.9 is an excerpt from the data dictionary that accompanies each data file. The data dictionary describes the contents and record layout of the public-use microdata file and identifies all data items in the order in which they appear on the file. The data dictionary is printed in the technical documentation for each file and is also available as a machine-readable text file (without page breaks) on the same tapes as the data files.

For most data items, the first line of each item description, as in figure 4.9, with a "D" at the left margin, gives you three pieces of information:

Data Name: The data name field is an eight-character field. It may be a mnemonic such as "WEEKSLK" or may include a four-digit number, as in "SC1054" in figure 4.9. (The letters SC refer to source codes. Each answer field on the questionnaire is given a source code. The source codes are used in entering data into the computer and in programming the file. Where a four-digit number appears, the item corresponds directly to an item on the questionnaire designated with the same four-digit number (see the middle column of questionnaire pages in the technical documentation). It is frequently useful to work with the questionnaire side by side with the data dictionary.

Size: The size of the data field tells how many characters the data item occupies. In this example, the data item SC1054 occupies two characters.

Begin: The next field gives the character location in which the data field begins. In this example, the data item SC1054 begins in location 2834.

Selected data items contain data dimensioned in an array, i.e., consisting of a series of subfields of equal size. For these data items, the term TABLE also appears on that first line, as in the example of WEEKSLK shown in figure 4.9.

The next few lines contain descriptive text and notes. For item SC1054, the text restates the question on which the data are based. In other cases, the range of valid values for a continuous variable may be shown. For example, the variable WSI-2032, "amount of pay before deductions," can have values ranging from -9 (not in universe) to 33332, the highest monthly dollar amount allowed by top-coding.[13] Not-in-universe codes are assigned if recipiency was reported to be "no" in the person segment.

---

[13]Monthly unearned income amounts (other than assets) are top-coded to $8,333 if the total amount for the wave exceeds $33,332. Hence, if an amount such as $20,000 was reported in reference month one and "0" reported for the other 3 months in one wave, no top-coding would have been instituted. However, if $20,000 was reported in each of the 4 months, then the monthly amounts are top-coded to $8,333. Asset income amounts are top-coded to $12,500 before construction of monthly amounts.

The line with "U" at the left margin designates the universe for the item; that is, the units or persons for whom information is provided. Following this, when necessary, are categorical value codes and labels. For example, for SC1054 there are 11 value codes with accompanying labels, beginning with "Ø." The first of these value lines begins with a "V" to distinguish them from the universe line or lines preceding the value labels.

The machine-readable version of the data dictionary may in some cases allow for direct input of format information into the user's software. Many of the major statistical or data management packages (e.g., SPSS, OSIRIS, CENSPAC) allow data dictionary information to be read in from an external file. Unfortunately, there are no standards for how this data dictionary information is recorded. Thus, in order to use the SIPP data dictionaries on tape, it is necessary to write a program to reformat the SIPP data dictionary as required for any given software. Alternately, if the number of data fields being used is relatively small, a text editor may be used.

The data dictionaries for core files, Waves 2 to 9, are nearly identical. Between Waves 1 and 2 there are a number of changes in location due to minor differences in the structure of the questionnaires and the inclusion of certain items. User notes are issued highlighting the changes in variable locations between the waves.

TECHNICAL CONVENTIONS OF SIPP TAPES

Some of the most important technical conventions associated with SIPP tapes are as follows:

Densities: The Census Bureau offers 6250 bpi and 1600 bpi on 9-track tape.

Recording Language: Users have the option of EBCDIC or ASCII.

Labels: EBCDIC tapes are given IBM labels; ASCII tapes carry ANSI labels. Users may request unlabelled tapes, though it is not recommended.

Block Size: SIPP records are typically packed together into blocks which are a multiple of the record size, but no more than 32,000 characters. Users may specify a smaller block size.

File size: Complex (or relational) files with core data typically require three high density (6250 bpi) tapes for one wave, while rectangular files typically require two, assuming a large block size. Topical modules are expected to be larger.

FIGURE 4.9 - EXCERPT FROM DATA DICTIONARY OF WAVE 1 RECTANGULAR FILE

PERSON

DATA	SIZE	BEGIN	
D WEEKSLK	18	2649	TABLE

Was this person looking for work or on layoff during this week of the reference period. There are 18 answer fields, one for each week of the reference period.

U Persons 15 years old or older

V 0. Not applicable
1. Yes
2. No

.
.
.

D SC1054 2 2834

During the weeks that ... wanted a job but was not looking for one, what was the main reason ... was not looking

U Persons 15 years old or older

V 0. Not in universe
1. Believes no work available in line of work or area - skip to SC1240
2. Couldn't find any work - skip to SC1240
3. Lacks necessary schooling, training, skills, or experience - skip to SC1240
4. Employers think too young or too old - skip to SC1240
5. Other personal handicap in finding job - skip to SC1240
6. Can't arrange child care - skip to SC1240
7. Family responsibilities - skip to SC1240
8. In school or other training - skip to SC1240
9. Ill health, physical disability - skip to SC1240
10. Other - skip to SC1240

.
.
.

D W51-2032 5 3298

What was the total amount of pay that...received before deductions on this job last month (month 4). Range = -9,33332

U Persons 15 years old and older

V -9. Not in universe
0. None

OTHER ASPECTS OF WORKING WITH SIPP FILES

The following chapters deal with further aspects of working with SIPP microdata files. Chapter 5 discusses how to use the public-use files for estimation. Topics include derivation of weighted estimates; use of partial samples; estimates involving households, families, and special program recipiency units; and estimates for subnational geographic areas.

Chapter 6 discusses the problems associated with matching data from more than one cross-sectional file. Chapter 7 deals with the calculation and use of standard errors from the SIPP sample, as well as other factors affecting the accuracy of the data.

REFERENCES

McMILLEN, D. and R. HERRIOT. "Toward a Longitudinal Definition of Households," Journal of Economic and Social Measurement (formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 349-360.

DUNCAN, G. and M. HILL. "Conceptions of Longitudinal Households: Fertile or Futile?," Journal of Economic and Social Measurement (formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 361-375.

CITRO, C., D. HERNANDEZ, and R. HERRIOT. "Longitudinal Household Concepts in SIPP: Preliminary Results," Second Annual Research Conference Proceedings, March 23-26, 1986. Washington, D.C.: U.S. Bureau of the Census, 1986.

CITRO, C., D. HERNANDEZ, and J. MOORMAN. "Longitudinal Household Concepts in SIPP," Proceedings of the Section on Social Statistics, American Statistical Association, 1986.

CHAPTER 5 - USE OF CROSS-SECTIONAL FILES FOR ESTIMATION AND ANALYSIS

SIPP public-use cross-sectional files are designed to permit analysis of the data in a manner which is not possible through the use of published reports. However, there are a number of characteristics of these files which make their use for cross-sectional estimation more difficult than is currently true for cross-sectional surveys like the CPS. These characteristics include:

- The cross-sectional SIPP files contain data collected as part of a longitudinal survey.
- Unlike true cross-sectional surveys, household and family composition is not constant throughout the reference period.
- The sample design is complex thus requiring the use of sample weights for each observation; weights corresponding to each month of the reference period are available.
- Persons may not be part of the sample for the entire reference period of the wave.
- The data cannot be viewed in a simple hierarchical fashion like the March CPS.
- The reference period does not cover the same calendar period for each observation.
- Less than the full sample may be available for most calendar-period estimates.
- Geographical information is not revealed entirely and in some cases is randomly changed to avoid disclosure.

This chapter elaborates on these characteristics which complicate use of the files for cross-sectional estimation and provides the user with instruction on how to prepare the desired estimates. It attempts to cover most of the general types of problems the user may encounter and includes examples for clarification. The chapter is organized into the following major sections:

- Identification system
- Person as the unit of analysis
- Households and families as units of analysis
- Other analytic units
- Preparing estimates for different periods of time
- Preparing estimates for different geographic units

IDENTIFICATION SYSTEM

To facilitate keeping track of respondents in the field as well as linking of the data across waves, each person is assigned an identification number when he or she enters the survey. The ID number consists of 14 digits identifying the

sampling unit and each person within it. There is also a series of two-digit fields referring to the person's address each month (current address ID) which changes when he or she moves. The 14-digit person ID provides a means of linking information about an individual from wave to wave. The two-digit current address identifiers uniquely identify the household of which each person is a member at any point in time. Thus it is possible to link data from all persons throughout the 2 1/2-year duration of a panel. This also allows the construction of monthly household income estimates based on the actual composition of households during the measurement period.

Identification Numbers

The person ID number consists of the following parts:

Sample Unit ID (SU-ID)	9 digits
Entry Address ID (PP-ENTRY)	2 digits
Person Number (PP-PNUM)	3 digits

The sample unit ID is a scrambled version of a census code for the respondent's sampling area, cluster of housing units within that area (called a "segment"), and a sequentially assigned serial number. With the exception of merged household members (described below), the sample unit ID never changes. It is the link to the original sample unit. All additional persons who are associated with an original sample person will receive the same sample unit ID.

The entry address ID represents the address of the person at the time he/she was first interviewed and does not change even if the person moves. It is used in conjunction with the person number to uniquely identify persons within the sample unit. This variable is the number 11 for all original sample members. For additional sample members, this can be 11 or greater than 11 depending on the current address ID of the unit which the new sample member joined.

The person number is a three-digit number, the first digit of which indicates the wave in which the person entered the survey. The remaining digits are assigned sequentially within the household. Original sample members are assigned person numbers in the 100 series. New sample persons are assigned person numbers in the N*100 series where N is the wave when they were first interviewed.[1]

The current address ID (H*ADDID--as listed in the technical documentation) together with the sample unit ID uniquely associates each person in the sample unit with an address in each month. A group of people sharing the same address in the same month within a sample unit is referred to as a household. The current address ID changes whenever an individual or group of individuals moves to a new address. The first digit of the current address number identifies the Wave (1 to 9) in which a household is first scheduled for interview at a new address; 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 a different location by original sample persons.

---

[1]The user should note that, while in many cases the household reference person is an original sample member and may be the first person in the household, this is not always true. Location of the reference person within household is discussed later in this chapter.

In Wave 1, all households have 11 as the current address number. In Wave 2, new households are numbered 21, 22, and so on. (Examples follow to illustrate how this scheme works.)

Thus far, several situations in which the ID system must be adjusted have been identified. For example, in Wave 2 of the 1984 Panel, two separate households previously selected into the sample merged by marriage. In this situation, one of the original set of ID numbers was retained for the merged household, and the new person numbers were assigned to the person who lost the original identifiers. Interviewers recorded new and old ID numbers to provide linkage and these are retained in the data files for the wave after the move. Individuals whose sample unit ID and person numbers have been changed are referred to as merged household members.

Merged household members are relatively rare and their existence can be ignored for cross-sectional analysis. However, if the user wishes to link public-use cross-sectional files over time, their existence poses some difficulties. These issues are discussed at length in chapter 6 which provides instruction on how to link public-use cross-sectional files over time.

Merged household members can be identified after the move takes place in several ways. First, these individuals will have person numbers in the range 180 to 189. Second, they will have previous wave identifiers retained in the following fields:

Old Sample Unit	PREV-ID
Old Entry Address	SC0064
Old Person Number	SC0066
Old Rotation Group	PREV-ROT

Finally, these individuals can be identified from the User Note Series issued to users who purchase the public-use tapes and associated technical documentation (available from Customer Services, Data User Services Division, Bureau of the Census, Washington, D.C. 20233 (301-763-4100)). In that series, the Census Bureau provides a list of all such persons using both the old and new identifiers.

Examples of Change in Sample Unit Composition

The following example illustrates how the ID system works as members of households move to new addresses, additional persons move in with them, and households split. The example shows how the ID system works for the hypothetical sample unit illustrated in figure 2.4.

Let us say that, in Wave 1, there is a five-person household in Detroit consisting of a husband, wife, daughter, son, and cousin (see figure 5.1). Since this is the first wave, the current address number is 11--indicating Wave 1, address 1--and the entry address number for each member of the household is the same as the current address number. Since they are assigned in Wave 1, the person numbers are in the 100 series and numbered sequentially, beginning with 101.

In Wave 2, the son leaves the sample because he joined the army. In Wave 3, the daughter marries and her husband moves into the household with his wife and in-laws. The current address number where the mother, father, cousin, daughter,

and son-in-law live remains the same since it is the same address. The son-in-law's entry address number is 11 since he first enters the survey at an address coded 11. The person number for the son-in-law is in the 300 series (301) since he joins the survey in Wave 3.

During Wave 4, the daughter and son-in-law move into a new house. The current address number of the daughter and son-in-law is 41 since they establish a new address in Wave 4.

The situation remains unchanged until Wave 7 except that in Wave 5 the cousin is taken by an uncle who is not in the sample, and, therefore is not followed. In Wave 7, the parents move to California. They retain the original person identifiers but are assigned a new value (71) for their current address denoting the move took place in Wave 7.

During Wave 8, the parent's household splits up. The wife moves to Florida while the husband moves to Detroit. The current address ID where the husband lives is 81 since he moved during Wave 8. His entry address ID and person number stay the same. The current address number where the wife now lives is 82, i.e., the second household added into the SIPP sample in the eighth wave. (The assignment of address number 81 to the husband and 82 to the wife is random. It could be the other way around.) The wife's entry address ID and person number stay the same.

While the parents were breaking up, the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. However, the grandchild's entry address is 41 instead of 11 because that is the current address ID of the daughter and son-in-law at the time of birth. The person number assigned is 801 reflecting the fact that the grandchild came into the sample in Wave 8.

PERSON AS THE UNIT OF ANALYSIS

While the data are issued as cross-sectional and cross-sectional estimation is the topic at hand, the longitudinal nature of the sample and survey design and the procedures employed by the Census Bureau in producing the public-use tapes complicate the calculation of cross-sectional statistics for persons. This section elaborates on the issues the user should consider in producing person-based estimates.

The data are collected and issued on a wave-by-wave basis. However, the majority of the cross-sectional applications are to produce monthly statistics. The manner in which people enter and leave the sample and the failure to interview adults affect the production of these monthly estimates. Persons can be present in the sample in all 4 reference months and the interview month, they can enter the sample after the first reference month of the wave, they can leave the sample after the first month of the wave, or they can become non-interviews. The classification of each respondent in these situations determines whether the individual should be included in monthly estimates for some months in the wave, all months, or not at all. This classification also indicates whether the monthly data were derived from an actual interview or imputed. Imputation, as discussed in this section, refers to the imputation of entire

FIGURE 5.1 - IDENTIFICATION SYSTEM FOR THE SAMPLE UNIT DEPICTED IN FIGURE 2.4

	Sample Unit	Current Address ID	Entry Address ID	Person Number
Wave 1				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 2				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Cousin	101111103	11	11	105
Wave 3				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son-in-Law	101111103	11	11	301
Cousin	101111103	11	11	105
Wave 4				
	Parent's Household			
Father	101111103	11	11	101
Mother	101111103	11	11	102
Cousin	101111103	11	11	105
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Wave 7				
	Parent's Household			
Father	101111103	71	11	101
Mother	101111103	71	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Wave 8				
	Father's Household			
Father	101111103	81	11	101
	Mother's Household			
Mother	101111103	82	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Grandchild	101111103	41	41	801

questionnaires in cases of noninterview. This type of imputation is accomplished using a two-step procedure. First, control card items are brought forward from the previous wave. Second, all questionnaire items are imputed from other interviewed individuals within the same wave using a statistical match procedure.[2]

Some important concepts to note:

- Month-in-sample (person) codes (PP-MIS) denote which month(s) the observation should be counted (PP-MIS-* = 1 implies the observation should be counted in Month *).

- Monthly demographic data derived from the control card are all set to zero in the months the person is not in sample, i.e., when PP-MIS-* = 2. The variables affected are

  Weight (FNLWGT-*)
  Relationship to reference person (RRP-*)
  Age (AGE-*)
  Marital status (MS-*)
  Family type (FAMTYP-*)
  Family relationship (FAMREL-*)
  Family number (FAMNUM-*)
  Person number of spouse (PNSP-*)
  Person number of parent (PNPT-*)

  Where * denotes the reference month (1, 2, 3, or 4) or the interview month (5). The user should always check the value for PP-MIS before using these demographic data.

- Information from the questionnaire is imputed when the person was "in sample" for at least one month of the reference period (PP-MIS-* = 1 for * = 1, 2, 3, or 4) but not actually interviewed (PP-MIS-5 = 2). Note this includes persons who moved out of scope as defined in chapter 2 and persons who died, as well as persons who refused to be interviewed in the wave. Other cases when the whole questionnaire was imputed are Type 2 noninterviews (PP-INTWV = 3 or 4).

- Data are collected for all months of the reference period even if the interviewed person was in the sample only for part of the reference period. The user should ignore this information for months for which PP-MIS-* = 2. Also, persons for whom the questionnaire has been imputed will have the data imputed for all months even though the person was not in the sample all months. Again the user should ignore the data in months for which PP-MIS-* = 2.

---

[2]Nonrespondents are matched to respondents based on demographic characteristics.

Figure 5.2 illustrates the various situations which can occur within waves, expanding the four general classes mentioned earlier into seven specific situations and adding a category for children for whom no interviews are conducted.

To a large extent this categorization is based on interview status, differentiating among various types of noninterview. The types of noninterview are defined in chapter 7.

The majority of people in each wave fall into Category A, interviewed original sample adults. These are people that the survey is designed to follow over time and for whom an interview was successfully obtained in the wave. Category B consists of interviewed adults who are included in the sample only for the period that they reside with an original sample adult (interviewed new sample adults).

Category C--merged household members, the third group of people in figure 5.2--are original sample members who moved into a household with other original sample members who were initially in a different sample unit. These cases are problematic for the linkage of two waves of data for two reasons: (1) the primary reason--the person identifiers change over time; (2) the secondary reason--the possibility of having two records for these merged household members within one wave. In fact, the only time there is only one record within a wave is when the move took place very early in the reference period.

When there are two records within a wave for a merged household member, one record will exist with the original identifying variables and the other with the new identifying variables (previous wave identifiers exist in this latter record containing the old identifiers). The former record will be treated like a Type D nonresponse in the data processing phase and, hence, all economic data will be imputed cross-sectionally through a statistical match procedure. The latter record will contain the results of the actual interview.[3] The month-in-sample codes (PP-MIS) on the file will be set so that the former record will be "in sample" prior to the move and the latter, "in sample" subsequent to the move. In each record, the weight fields and edited monthly demographic data will be nonzero only in the months considered "in sample."[4] In the case where a merged household member only has one record in the wave, it will contain the new identifiers and all data relevant to that wave.

Categories D, E, and F--the fourth through seventh categories in figure 5.2--are different types of cases where no interview was conducted in the wave but a record for that person exists on the file. Category D (Type Z noninterview) consists of original sample adults in a household where other adults responded to the survey.

---

[3]Note that this could also represent a Type Z nonrespondent, in which case the data are imputed in both records.

[4]There is a perfect correlation between being "in sample" and having positive weights and valid monthly demographic data.

FIGURE 5.2 - SOURCE OF QUESTIONNAIRE INFORMATION: REPORTED VS. IMPUTED

Category	Months In Sample	Source For Data 1/	Description
A	All	Interview	Interviewed original sample adult
B	All	Interview	Interviewed new sample adult, i.e., a person with person number in the 200+ category.
	Some	Interview	
	Some	Imputed	
C	Some or All	Interview	Merged household member whose basic identifiers changed (merged households are discussed in chapter 4.)
	Some	Imputed	
D	All	Imputed	Nonrespondent with imputed data who was considered present in the sample the full reference period (Type Z nonresponse)2/.
E	Some	Imputed	Mover who was not successfully interviewed and hence the data were imputed. They were not considered present in the sample after the move (Type D nonresponse and selected Type A noninterview3/.)
F	Some	Imputed	Original or new sample member who left the sample because of the design (i.e., death, moved out of scope, new sample member no longer residing with original sample member). Data were imputed and the person was present only until the month he/she left.
G	None	None	Other nonrespondents for whom there were records buy imputations were not performed. The person was not considered in sample for the entire wave (Type C noninterview and Type A noninterview not included in Category E.)
H	All	Control	Children under the age of 15. The length of time they are in sample
	Some	Card	is a function of the adults with
	None		whom they reside.

1/If the source is "imputed," then the demographic data from the control card is obtained from the previous interview and the questionnaire items are imputed through a statistical match procedure.

2/In Wave 1 of the 1984 Panel, interviewed households containing one or more Type Z nonrespondents were assigned zero weights. Data for Type Z nonrespondents in this case were not imputed. After Wave 1 of the 1984 Panel, interviews of households containing one or more Type Z nonrespondents were assigned positive weights.

3/Members of Type A noninterview fall in this category if they resided with an interviewed sample adult early in the reference period and, hence, were classified as "in sample" for at least one month in the wave.

After Wave 1 of the 1984 Panel, Type Z noninterview adults are treated as present in the sample the full wave (PP-MIS codes are set to 1 for all months). However, the data from the questionnaire are imputed cross-sectionally through a statistical match procedure. In Wave 1 of the 1984 Panel, Type Z noninterview adults and all other members of their Wave 1 household were assigned zero weights.

Category E includes cases where an original sample member or a group of original sample members moved within the wave but could not be located for interview (Type D noninterview) or refused to continue participation (selected Type A noninterview).[5] These cases are treated like Type Z nonresponse (Category D) except that they are only considered "in sample" for the months prior to the move. (PP-MIS codes switch from 1 to 2 at the point the move occurs.) The weight fields and monthly edited demographic variables are only nonzero for the months the person is considered as "in sample." This category is distinguished from Category G (described below) in that persons in Category E resided with an interviewed adult during at least one month of the reference period. Note that the months these movers were "in sample" are determined by the date they were reported by the interviewed adult to have left their former household.

Category E can be illustrated by an example. A husband/wife family with child was successfully interviewed in Waves 1 and 2. At the beginning of the third month of Wave 3, the wife and child move but are not immediately located. In Wave 3, the husband is interviewed, but the wife refuses resulting in a Type A noninterview. Because the husband reports the wife and child residing at his address in month 1 and 2, the wife and child are flagged as "in sample" during those 2 months. Due to the noninterview situation, however, questionnaires were not administered. As a result, the questionnaire information for the wave is imputed so that data exist on the files for the months the wife was reported to be in the sample.

Category F consists of original sample adults who left the sample due to death, institutionalization, entering military barracks, or a move outside the geographic areas where an interview could be conducted. These individuals are treated like the movers described in Category E. That is, since no interview was conducted, the majority of the data are imputed. Also, month-in-sample codes are set to "1" for before the individual left and changed to "2" for after, with the corresponding suppression of weights and edited monthly demographic information.

Category G consists of those original sample adults in households for which no interview was successfully conducted. In these cases, records exist for the persons in the household but no data were collected and no imputations were performed to compensate for the lack of information. Individuals falling into this category consist of those in Type C noninterviewed households, as well as those in Type A noninterviewed households not previously classified as Category E.

---

[5]Members of Type A noninterview fall in this category if they resided with an interviewed sample adult early in the reference period and, hence, were classified as "in sample" for at least one month in the wave.

The final category in figure 5.2, Category H, consists of children under the age of 15. The survey only includes a provision to collect minimal data for these individuals and they are not followed unless they reside with an original sample adult. Children have their "in sample" codes set based on the codes for the other adults in the household at the time of the interview. As is true for adults, the weight fields and edited demographic data are set to zero in months classified as "not in sample."

Use of Weights

The SIPP was originally designed to obtain a self-weighting probability sample. In a self-weighting sample, every sample unit has the same overall probability of selection. For the 1984 Panel, in self-representing PSU's, the sampling rate is about 1 in 3,700; in nonself-representing PSU's, the sampling rate is higher, as the sampling is adjusted to account for the PSU's probability of selection. For example, if a nonself-representing PSU was selected with a probability of 1/10, the sampling rate within the PSU would be roughly 1 in 370 instead of 1 in 3,700. Because of changes in the sample, SIPP is no longer self-weighting. Also weights used in the preparation of SIPP estimates have been adjusted in order to improve the precision and accuracy of these estimates.

The estimation procedure used to derive SIPP person weights involves several stages of weight adjustments. In the first wave, each person received a base weight equal to the inverse of the probability of his/her address being selected. In subsequent waves, each person received a base weight that accounted for differences in the probability of selection caused by the following of movers to new addresses (Huang, 1984).

A noninterview adjustment factor is applied to the weight of each interviewed person to account for persons in occupied living quarters who were eligible for the sample but were not interviewed (see Kalton and Kasprzyk (1986), and Chapman, Bailey, and Kasprzyk (1986)). (See chapter 7 for more about noninterviews.) A factor termed "Stage 1 adjustment factor" is applied to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata from which they were selected.

A Stage 2 adjustment to persons' weights is performed to bring the sample estimates into agreement with independent monthly estimates of the civilian (and some military) noninstitutional population of the United States by age, race, and sex. These independent estimates, described in "Population Estimates and Projections" (Current Population Reports, Series P-25, Number 985), are based on statistics on births, deaths, immigration and emigration, and the strength of the Armed Forces. To reduce sample error, weights are further adjusted in such a manner that SIPP sample estimates would closely agree with Current Population Survey (CPS) estimates by type of householder (married, single with relatives, or single without relatives, by sex and race) and relationship to householder (spouse or other). The estimation procedure also involved an adjustment that was needed to make the husband and wife of a household receive the same weight.[6] Starting in the fifth wave of the 1984 Panel,

[6]The need for this last adjustment results from the differential treatment of males and females in the prior adjustments.

a further adjustment is made to base weights to compensate for the subsampling, made necessary by budgetary problems.

This Stage 2 adjustment procedure resulted in the assignment of five separate weight fields for each person in each wave. The five fields correspond to the 4 reference months and the interview month. Users should use a different weight field for each set of monthly cross-sectional estimates derived from a wave.

The weight estimation procedure described above resulted in persons' weights varying from about 400 to about 40,000, although this range varies by wave. Figure 5.3 displays basic statistical properties of the weights in the first five waves of the 1984 Panel. After Wave 1, persons in the sample for less than the entire 4-month period received zero weights for months not in the sample, i.e., when PP-MIS-* = 2. Most statistical software packages handle weighted data with no difficulty. In tabulating a characteristic, the software can apply the appropriate weight to each response. Different software packages handle weighting with varying degrees of efficiency, however.

Figure 5.4 illustrates the fact that the use or nonuse of weights can significantly affect the outcome of tabulations. Three of five persons in the illustration work full time; two do not. But, since the persons who do not work full time happen to have higher weights than the others, weighted totals show the two groups to be equal.

Individual Characteristics

For the most part, the economic and demographic data for individuals can be used directly, as is true for the public-use data files from any other household survey. The exceptions are as follows:

- Ignore all monthly data in months for which PP-MIS-* = 2. These are months for which zero weights have been assigned.

- On the rectangular files (except in Wave 1 of the 1984 Panel), nonreceipt of a particular income source for an entire wave results in the assignment of "9" to the monthly recipiency flags and "-9" to the monthly benefit amounts. Users should take care not to treat "-9" as a valid dollar amount. In Wave 1, recipiency flags are set to "2" and amounts are set to "0."

Users wishing to distinguish imputed or edited data from reported data should check the following:

- If PP-MIS-5 = 2 and PP-MIS-* = 1 for at least one reference month or if the person is Type Z noninterview (PP-INTWV = 3 or 4), then the entire questionnaire has been imputed through a statistical match procedure and the control card data have been brought forward from the preceding wave.

FIGURE 5.3 - STATISTICAL PROPERTIES OF PERSON WEIGHTS ON 5 WAVES OF THE 1984 PANEL

Reference Month	Number of People with Positive Weights	Minimum Weight	Maximum Weight	Mean	Standard Deviation
Wave 1					
1	52758	1418.2906	23893.560	4369.5332	1119.5720
2	52758	1499.3005	23832.698	4373.4528	1117.2093
3	52758	1493.3164	23938.858	4377.3246	1118.7028
4	52758	1521.9843	23340.086	4380.8124	1117.1528
5	52758	1574.8917	22697.291	4385.1969	1122.9979
Wave 2					
1	38347	652.6406	23109.877	4522.4067	1172.1690
2	38605	596.9902	22756.618	4496.3410	1198.4732
3	38891	587.9985	22102.665	4467.2325	1225.4372
4	39249	574.3549	22629.997	4430.8427	1253.6847
5	39463	376.7760	22802.065	4409.4225	1269.7611
Wave 3					
1	50855	373.1103	25650.175	4560.4766	1268.2240
2	51058	570.6232	23442.499	4545.5409	1282.2946
3	51326	510.6788	23260.318	4525.4225	1302.3464
4	51652	464.0937	23108.967	4500.5074	1310.0838
5	51956	437.5975	23541.411	4477.6631	1330.5957
Wave 4					
1	49725	437.8826	25151.575	4677.2763	1395.0946
2	49931	447.2045	25962.227	4662.0554	1405.1208
3	50181	418.1458	30636.813	4642.7362	1419.3529
4	50488	414.6399	29189.175	4618.8259	1434.5697
5	50929	486.0649	27755.260	4581.8922	1436.5818
Wave 5					
1	45377	459.4248	37183.222	5142.2325	1764.6304
2	45504	442.0941	35393.756	5131.4182	1766.6433
3	45669	487.5410	33310.011	5117.2487	1770.8800
4	45842	495.7347	33636.873	5101.8817	1782.5999
5	46063	493.4739	33180.704	5081.4399	1783.0478

FIGURE 5.4 - EXAMPLE OF WEIGHTED DATA

	Worked Full Time	Weight	Raw Counts No	Raw Counts Yes	Weighted Counts No	Weighted Counts Yes
Person 1	No	4,000	1		4,000	
Person 2	No	5,000	1		5,000	
Person 3	Yes	3,000		1		3,000
Person 4	Yes	3,000		1		3,000
Person 5	Yes	3,000		1		3,000
			2	3	9,000	9,000

- Control card information is repeated twice, once in unedited fields reflecting the reported amount and once in edited fields. A comparison of the two will denote when the original reported amount was changed.

- Item imputation is denoted by separate flags for each variable imputed.

The various forms of imputation and the use of the imputation flags are discussed more fully in chapter 7.

Household and Family Characteristics

In each month, persons considered "in sample" can be arrayed by the characteristics of the households and families to which they belong in each month. On the public-use tapes, the Census Bureau has generated summary characteristics of these units which can be used directly in tabulations of persons.

If the existing summary fields are not sufficient, the user can create new summary characteristics as desired by summing the relevant variables for all individuals in the same unit in the same month. Grouping persons in households or families with SIPP files is more complicated than with files in which person records are arranged in a strictly hierarchical fashion within household. As discussed in chapter 4, address records in SIPP complex files carry pointers to each person who was a resident at the address in each month. There are five sets of pointers, one for each month of the reference period and one for the interview month. The analyst would need to group persons in households using these address identifiers.

The rectangular file does not have these house-to-person pointers within sample unit, but does identify the current address ID (H*-ADDID) of the household of which the person was a member each month. The file can be sorted on current address ID within month and within sample unit to group household members together for any particular reference month. Another option available to rectangular file users is to sort on the person number of the householder within month and within sample unit, which is provided on each household member's record. Note, however, that, if estimates are being derived for more than one month for each person record, a separate sort is required for each month because composition can change from one month to the next. Further information on forming household and family units is provided later in the discussion of these as units of analysis.

In addition to characteristics of the unit to which the person belongs each month, some indication of the stability of that unit within the wave can be derived from the household and family change fields (HCHANGE*, FCHANGE*, and SCHANGE*). Longitudinal units have not been defined on the public-use cross-sectional files. However, the change fields will indicate if any change at all occurred in 2 successive months whether it be a change in household composition or an address change by the entire unit. This change may disrupt the unit (such as a divorce) or it may not (such as the birth of a baby). Household change fields also denote a change of address as well as a change in composition. Hence, a unit which moves intact to a new address will have change fields set to denote a change took place. The change fields cannot be used except by comparing two successive values in the following manner:

Month 1 - Cannot tell if a change occurred from the previous wave.

Month 2, 3, 4 - If the change field in one month equals the change field in the prior month, no change took place. Otherwise, some change took place.

Month 5 - Change only defined for households in the interview month. The field is used as noted for months 2, 3, and 4.

Note specifically that the value in any one field has no meaning.

Top-coding of Income Variables

To protect against the possibility that a user might recognize the identity of a SIPP respondent with very high income, income from every source is "top-coded" so that no individual annual income amounts above $100,000 per source are revealed. While the data dictionary indicates a top-code of $33,332 for monthly income, this top-code will rarely be used. In most cases the highest monthly income is shown as an individual dollar amount of $8,333, with $8,333 actually representing "$8,333 or more." (The $100,000 annual income top-code is $8,333 multiplied by 12 months.) Individual monthly amounts above $8,333 may occasionally be shown if the respondent's income varied considerably from month to month, as long as the total amount reported in the wave did not exceed $33,332. For example, if a respondent's income from a single job were concentrated in only 1 of the 4 reference months, a figure as high as $33,332 could be shown.

There are exceptions to the general rule for top-coding amount fields. In particular, income from interest or property is modified so that the total amount recorded per wave does not exceed $12,500. Similarly tuition and fees on the person records have been modified so as not to reveal more than $4,000 per wave. Hourly rate of pay (SC2028) is top-coded at $52.00.

Top-coding is performed for the individual income types on each person's record. To derive one person's total income, the top-coded individual income types are aggregated. To derive total family income or total household income, the top-coded individual income types are aggregated for each member of the family or the household, respectively. Thus, a person with high income from several sources (jobs, businesses, property) could have aggregate monthly income well over the top-code for each source. Families and households with a number of high income members could theoretically have aggregate income well above $100,000, though well below the $1.5 million shown as the highest allowable value in the data dictionary.

The user is cautioned against trying to make much use of the occasional monthly figures above $8,333, except in calculating aggregates or observing patterns across the 4-month period for a single individual, family, or household. Those units with higher total monthly amounts shown are a biased sample of high income units, more likely to include units with income from multiple sources rather than units with equally high aggregate income which comes from a single source.

HOUSEHOLDS AND FAMILIES AS UNITS OF ANALYSIS

Cross-sectional estimates of the characteristics of households and families in each month can be derived from the public-use cross-sectional files. Longitudinal analysis of households and families cannot be conducted unless the user first imposes a longitudinal unit definition. Refer to McMillen and Herriot (1985) and Duncan and Hill (1985) for a discussion of the issues surrounding the longitudinal definition of households and families. Note that an address identifier is used to define monthly households and this identifier is often the same from one month to the next. However, the user should not use this variable to link households across months because it only denotes a common dwelling unit. Intact households can change addresses (and, hence, have different address identifiers), and units with common addresses across 2 consecutive months could have experienced a substantial change in composition.

Issues to consider in using the public-use files to produce monthly estimates of families and households are described below.

Formation of the Unit

The public-use files contain selected characteristics of monthly households and families which can be used directly in the analysis. On the complex file, this is accomplished by tabulating households in the relevant month which have HH-MIS = 1 from the address record or tabulating families in the relevant month from the family month record. On the rectangular file, the user should be careful to avoid double-counting households and families because these data are repeated on all person records. To avoid double-counting, the user can select one person record per household or per family in a specific month by screening out every one except the reference person of the unit. The record of the reference person of the household in the month is the record in which the following condition is met in the month of interest:

1,000 * PP-ENTRY + PP-PNUM = H*REFPER

The record of the family reference person is the one in which the following condition is met within the month of interest:

1,000 * PP-ENTRY + PP-PNUM = F*REFPER

The record of the subfamily reference person is the one in which the following condition is met within the month of interest:

1,000 * PP-ENTRY + PP-PNUM = S*REFPER

It is possible that the data needs for a particular study may require the user to construct characteristics of households or families not already created by the Census Bureau. As discussed under the previous section on persons as the unit of analysis, monthly household and family characteristics can be constructed from the public-use files. The units are first created using the monthly current address identifiers and family numbers and in the case of the complex files, a series of monthly pointers. The desired unit characteristics are then created by summing over the records of the unit members. The user should note

that households and families may not exist for all 4 months of the wave. As an example, refer back to figure 5.1. Note that between Waves 3 and 4, a new household was formed. This actually occurred when the daughter and son-in-law moved out in the third month of Wave 4. Household 41 only existed in months 3 and 4. Household 11 existed for all 4 months because the parents did not move. Hence, in preparing monthly estimates of households, one household is counted in months 1 and 2 and two households are counted in months 3 and 4.

In creating summary economic characteristics of the unit, the user should pay particular attention to one inconsistency existing in the data files. Except for Women, Infants and Children Nutrition Program (WIC) vouchers, all income and benefits are recorded exclusively in the adult records. Therefore, some computational costs can be saved by restricting the summation to adults. However, this is not true for WIC vouchers. Amounts imputed for young children are actually assigned to the records of young children. Hence, these records must be included if WIC benefits are to be included in the household totals.

To determine household and family composition each month, the user must rely on the month-in-sample codes for each individual, as well as the unit identifiers and (in the case of the complex file) pointers. For example, figure 5.5 illustrates the composition of the hypothetical sample unit in Wave 8. In months 1 and 2, there are two households, each with two people. In the third month, there are three households--the first two contain one person each and the third, two. In the last month of the reference period, the third household is increased to three after the grandchild is born. Note that there will be a full record for the grandchild, but the grandchild should be ignored in the months before its birth, i.e., when PP-MIS-* = 2.

The user should note that households and families can be composed entirely of persons with PP-MIS = 2. For example, a Type A noninterview address (described in chapter 7) will have all persons flagged as "not in sample" in the interview month (and possibly one or more reference months). For months when this occurs, affected households and families should be screened out when producing monthly estimates because all members of the unit will have a zero weight. This situation occurs because the cross-sectional files contain records for all persons that the Census Bureau attempted to interview. However, when all attempts to interview at an address failed, other persons in the sample were reweighted to compensate for the missing information. This is discussed in more detail in chapter 7.

The user should also note that residents at a noninterview address could be flagged as "in sample" in the early months of the reference period. In this case, they will be counted in forming households and families in the early months of the wave but not the later months. This situation occurs when residents of a common address split up in the middle of a wave, and only one of the two (or more) descendent addresses is successfully interviewed. All persons in this case are flagged as "in sample" before the move. After the move, only the residents of the successfully interviewed address are flagged as "in sample."

Weights

As is true for cross-sectional analysis of persons, weights must be used in the preparation of estimates of monthly households and families because SIPP

is not a self-weighting sample. On the complex file, weights for households and families are carried, respectively, on the address and family month records. There are separate weight fields for each month. The weighting process defines the weight of the household to be the same as the weight of the household reference person (householder), and the weight of a family or subfamily is equal to the weight of the family or subfamily reference person. On the rectangular file, where household, family, and subfamily segments appear on each person record, all of the applicable monthly weights can be found on the respective segments of that record.

FIGURE 5.5 - COMPOSITION FOR THE HYPOTHETICAL SAMPLE UNIT IN WAVE B

	Month				
	1	2	3	4	Interview
Father - Current Address	71	71	81	81	81
PP-MIS	1	1	1	1	1
Mother - Current Address	71	71	82	82	82
PP-MIS	1	1	1	1	1
Daughter - Current Address	41	41	41	41	41
PP-MIS	1	1	1	1	1
Son-in-Law - Current Address	41	41	41	41	41
PP-MIS	1	1	1	1	1
Grandchild - Current Address	-	-	-	41	41
PP-MIS	2	2	2	1	1
Number of Households	2	2	3	3	3
Persons in Household					
1st HH	2	2	1	1	1
2nd HH	2	2	1	1	1
3rd HH			2	3	3

Group Quarters

Selected addresses in each wave of the SIPP sample are composed entirely of unrelated individuals, none of whom is designated as the reference person of the unit. These situations are called group quarters and reflect noninstitutional living arrangements such as rooming houses. These units are not households in the strict sense of the term and often are screened in the production of household estimates. Group quarters units can be determined based on the relationship to reference person variables (RRP-*, * = 1,2,3,4,5) for the person whose identifier is used in the pointer to the reference person. That is, a group quarters unit in month "*" is one in which one person meets the following criteria:

H*REFPER = 1000*PP-ENTRY + PP-PNUM

and

RRP-* = 7

Top-coding

As discussed earlier, individual income sources on each person's record are capped so as not to reveal high incomes which are relatively rare occurrences. Summary amount fields in the address and family month segments are simple sums of components shown in the person records after top-coding and are not independently top-coded.

OTHER ANALYTIC UNITS

Some analyses involve summarizing data for units other than households or families. The SIPP core module contains sufficient information to allow the user to choose program units as a unit of analysis for participants in the following programs:

Medicare
Medicaid
Aid to Families with Dependent Children (AFDC)
General Assistance
Health Insurance
Railroad Retirement
Social Security
Veterans Compensation and Pensions
Food Stamp
Women, Infants and Children Nutrition Program (WIC)

For all of these programs, SIPP records for each adult and child contain coverage fields denoting whether the individual received benefits from the program in question in each month. Persons can be flagged as being covered if they received benefits directly (such as the authorized person to receive food stamps) or indirectly (such as a dependent of the authorized recipient of food stamps). These coverage fields allow the analyst to distinguish situations in which the program unit is a subset of the family or household unit. For example, Carlson and Dalrymple (1986) found that 18 percent of households receiving food stamps in September 1983 contained at least one member who was not part of the food stamp unit, i.e., did not benefit at all from food stamps. This estimate was derived using the monthly coverage fields for food stamps (FOODSTP*, *=1,2,3,4) in Wave 1 of the 1984 Panel. The procedure employed by Carlson and Dalrymple involved several steps. First, persons were grouped into households, as described earlier. Second, households with positive food stamp benefits (H*-FDSTP > 0 for the relevant month) were flagged as participants. Next, the coverage fields for each person in the participant households were examined. If at least one person was flagged as not being covered, that household was counted as having a program unit that was a subset of the household.

Analysts can perform similar analyses with the coverage fields for other programs. These fields are

Medicare	CARECOV*
Medicaid	CAIDCOV*
WIC	WICCOV*
Health Insurance	HIMNTH*
Veteran	VETS*
AFDC	AFDC*

General Assistance	GENASST*
Foster Care	FOSTKID*
Other Welfare	OTHWELF*
Social Security	SOCSEC*
Railroad Retirement	RAILRD*

Except for the WIC, these coverage fields were derived directly from the information reported to the SIPP interviewer. On the first two SIPP panels, WIC coverage was imputed to children under 6 if a mother reported participation in the WIC program. Beginning with the 1986 Panel, SIPP directly measures coverage under the WIC program for all persons.

In addition to examining whether members of participant households or families are covered under the programs listed above, SIPP provides information to permit the construction of program units (except for WIC units) to be used as a basis for analysis, instead of households, families, or persons. The questionnaire is designed so that one person in each program unit is deemed a primary recipient, and that person reports who else benefits from the program and (where applicable) the actual benefit received for the group. (Note there is no actual benefit collected for the health insurance programs.) The SIPP data files contain the list of person numbers of persons covered by each program on the primary recipient's record. That list of person numbers can be matched against the person numbers of other members of the household in each month to determine who is in the primary recipient's unit. Note that, if there is only one primary recipient for a particular program in the household in a given month, the coverage fields can be used more directly to determine the program unit. However, if there is more than one primary recipient and hence more than one unit, the coverage fields cannot be used by themselves to determine program units. Once the unit is constructed in a particular month, the weight for the unit should be assigned as the weight of the primary recipient of the unit.

To illustrate the construction of program units, consider the example in figure 5.6. This sample unit consists of one household each month with a primary family (persons 101, 102, 104, and 105), a subfamily (person 104 and 105), and an unrelated individual (103). The first two people in the household in each month are elderly and hence are entitled to be a separate food stamp unit. The first person reports receiving food stamp benefits in all months and that not everyone in the address is covered by his benefit (SC3100 denotes this information; source codes (SC) are item numbers assigned in the questionnaire). Since not everyone is covered, person 101 lists that he (person 101) and his wife (person 102) are covered by the $60 food stamp coupons he gets each month. (The fields SC3102 through SC3120 allow the primary recipient to record up to 10 people covered by the benefit. If there are fewer than 10 people listed, zeros are assigned to the unused fields.) Both he and his wife have the food stamp coverage fields set to 1 for all months, but only he has valid data for the other fields since he was deemed to be the primary recipient in the husband-wife unit.

Note this figure displays data as they appear in the rectangular file and, hence, all the food stamp fields contain values for the wife even though she was not a primary recipient. She is assigned -9 in the amount fields and zeros in the other fields. The complex file is substantively the same. However, the variables pertaining to the benefit amount and the list of covered persons are in the G1 records; the wife will not have a G1 record for food stamps because she did not report the benefit.

FIGURE 5.6 - MULTIPLE FOOD STAMP UNIT

SUID	CURRENT ADDRESS	ENTRY ADDRESS	PERSON NUMBER	FAMILY NUMBER	SUBFAMILY NUMBER	FOOD STAMP BENEFIT	ALL PEOPLE COVERED[1/] *(SC3100)	PERSON NUMBER OF PERSONS COVERED[2/] *(SC3102-SC3120)	FS COVERAGE FIELD[3/]
2222222222222222	11 11 11 11	11	101	1 4 7 10	0 0 0 0	60 60 60 60	2	101 102 0 0 0 0 0 0 0 0	1 1 1 1
2222222222222222	11 11 11 11	11	102	1 4 7 10	0 0 0 0	-9 -9 -9 -9	0	0 0 0 0 0 0 0 0 0 0	1 1 1 1
2222222222222222	11 11 11 11	11	103	2 5 8 11	0 0 0 0	-9 -9 -9 -9	0	0 0 0 0 0 0 0 0 0 0	2 2 2 2
2222222222222222	11 11 11 11	11	104	1 4 7 12	3 6 9 10	0 0 30 30	2	104 105 0 0 0 0 0 0 0 0	2 2 1 1
2222222222222222	11 11 11 11	11	105	1 4 7 12	3 6 9 10	-9 -9 -9 -9	0	0 0 0 0 0 0 0 0 0 0	2 2 1 1

1/This column shows whether all the persons in that unit were covered by food stamps (1 = yes, 2 = no, 0 = not in universe).
2/If fewer than 10 people listed, zeros are assigned to the unused fields.
3/1 = yes, 2 = no
*SC = source code. These codes are the item numbers of the questions in the questionnaires. (They are in multiples of 2.)

The third person in the household each month was not covered under anyone else's food stamp benefit nor did he report receiving benefits on his own. Therefore, his coverage fields are set to 2 for all months and the food stamp benefit and related fields are set to the appropriate "not applicable" codes.

The fourth person in the household each month reported receiving food stamps in the third and fourth months of the wave. She lists that her benefit covered herself (person 104) and her child (person 105). She and her child are, therefore, flagged as being covered in months 3 and 4. She also has nonzero benefits in those 2 months and zeros assigned when no benefit was received. The child is assigned the appropriate "not applicable" codes for the benefits and associated data.

If the user were interested in knowing who in this example was not covered under food stamps, an examination of the coverage fields would yield the fact that three persons were not covered in months 1 and 2 and one person was not covered in months 3 and 4. A tabulation of the number of persons with positive benefits within households within months would yield the fact that this sample unit had one food stamp unit in months 1 and 2 and two food stamp units in months 3 and 4. Finally, matching the list of persons covered on each primary recipient's record to the list of persons in the household each month would yield one food stamp unit in each of months 1 and 2, consisting of persons 101 and 102, and two food stamp units in months 3 and 4. The first unit in each month consists of persons 101 and 102 and the second unit in those 2 months consists of persons 104 and 105. Person 103 would not be included at all because he was never covered by food stamps.

To reiterate, analysis of WIC units does not follow the preceding guide although other means-tested cash tranfers do. WIC primary recipients are covered persons over the age of 6. WIC vouchers are assigned to adults and children, both of whom must be included in summing total WIC benefits.

PREPARING ESTIMATES FOR DIFFERENT PERIODS OF TIME

Each person and household is assigned five weights on each interview file, one for each of the 4 reference months and one for the interview month. Families and subfamilies are assigned only four weights since there is no attempt to define families as of the interview month. The four sets of reference month weights can be used to form reference month estimates. Reference month estimates can be averaged, however, to form estimates of monthly averages over some longer period of time. For example, to estimate the average monthly number of persons in a specified income range over the 4-month period, one would average the four reference month weights together, excluding months when an individual was assigned a zero weight.

The fifth weight is specific to the interview month. This weight can be used to form person or household estimates that specifically refer to characteristics as of the interview month. For example, one might want to estimate the number of unmarried adults living with an aged parent as of the latest observation. Interview weights can also be used to form estimates referring to the time period including the interview month and 4 previous months, although this is not recommended because characteristics as of the interview date may not reflect the entire reference period, i.e., persons could move, marry, or die during the reference period.

The interview weight can be used for estimating a few of the demographic characteristics, such as race or sex, and other information that appear on the file for the 4-month reference period as a whole, but not for each month.

None of these weights has been designed to yield the best estimates for a person's status over 2 or more months. This would be important, for example, in estimating the number of persons existing in October 1983 who experienced a 50-percent increase in income between July and August.

Calendar-Month and Quarterly Data

In tabulating SIPP data for a particular calendar month, one must keep in mind the survey design. Most waves include four rotation groups, interviewed in 4 successive months. Figure 5.7 is a schematic diagram of the 1984 Panel design.

Months, quarters, and years are shown along the top. Each cell shows the wave and rotation groups for which data are collected for each month. Thus, in the first interview, conducted in October 1983, data were collected from Wave 1-Rotation Group 1 households for the months of June, July, August, and September.

As successive rotation groups are interviewed, the 4-month reference periods advance by 1 month. Wave 1-Rotation Group 2 households were interviewed in November 1983 to obtain data for July through October. Wave 1-Rotation Group 3 households were interviewed in December for August to November data. September to December data were obtained from Rotation Group 4 in their January 1984 interview.

In deriving calendar-month or quarterly estimates from the cross-sectional wave files, it is important to know how many rotation groups were interviewed, as less than the full sample may be available for the calendar month or the quarter from a single wave of interviews. If this is the case, the estimates must be inflated by an appropriate factor to represent the full sample.

In some months, a full sample of four rotation groups from the same wave will be available. For Wave 1 (see figure 5.7), data for September 1983 were collected from the full sample. These data consist of month 4 for Rotation Group 1, month 3 for Rotation Group 2, month 2 for Rotation Group 3, and month 1 for Rotation Group 4. All of these figures (with appropriate weights) must be added together because any one rotation group includes only one-fourth of the SIPP sample.

In deriving calendar quarter estimates, a full sample consists of data for four rotation groups for each of the 3 months in the quarter. This would entail using data from two or three waves. For example, the fourth quarter of 1983 includes Rotation Groups 2, 3, and 4 from Wave 1 and Rotation Groups 1, 2, and 3 from Wave 2. Weighted data from all these rotation groups must be added together to form a full sample.

Note, however, that a full sample is not available for the third quarter of 1983. Or, for subsequent quarters, the analyst may not want to wait for another wave of data to become available. Procedures to use in deriving estimates based on a partial sample are explained below.

FIGURE 5.7 - 1984 SIPP PANEL REFERENCE PERIODS

Wave	Rotation Group	Interview Month	Reference Period: 1983 2nd Qtr	1983 3rd Qtr	1983 4th Qtr	1984 1st Qtr	1984 2nd Qtr	1984 3rd Qtr	1984 4th Qtr	1985 1st Qtr	1985 2nd Qtr	1985 3rd Qtr	1985 4th Qtr	1986 1st Qtr	1986 2nd Qtr
	1	Oct	J	JAS											
1	2	Nov		JAS	O										
	3	Dec		AS	ON										
	4	Jan		S	OND										
	1	Feb			OND	J									
2	2	Mar			ND	JF									
	3	Apr			D	JFM									
	4	May				JFM	A								
3	1	Jun				FM	AM								
	2	Jul				M	AMJ								
	3	Aug					AMJ	J							
	4	Sep					MJ	JA							
4	1	Oct					J	JAS							
	2	Nov						JAS	O						
	3	Dec						AS	ON						
	4	Jan						S	OND						
5	1	Feb							OND	J					
	2	Mar							ND	JF					
	3	Apr							D	JFM					
	4	May								JFM	A				
6	1	Jun								FM	AM				
	2	Jul								M	AMJ				
	3	Aug									AMJ	J			
	4	Sep									MJ	JA			
7	1	Oct									J	JAS			
	2	Nov										JAS	O		
	3	Dec										AS	ON		
	4	Jan										S	OND		
8	1	Feb											OND	J	
	2	Mar											ND	JF	
	3	Apr											D	JFM	
9	4	May												JFM	A
	1	Jun												FM	AM
	2	Jul												M	AMJ

Note that Waves 2 and 8 only contain three rotation groups. For subsequent panels, the figure would be similar except each panel begins in February of each panel year, the rotation groups are numbered differently, only eight waves are conducted, and only one wave contains less than the full sample (Wave 2 in the 1985 Panel and Wave 4 in subsequent panels).

Working With Less Than the Full Sample

Figure 5.7 shows that, for October 1983, data were collected from only three rotation groups of Wave 1. Thus, the sample size available is three-fourths that available for September. The preferred way to handle this is to acquire Wave 2, as well, and combine October data for Wave 2-Rotation Group 1 with the Wave 1 October data for Rotation Groups 2, 3, and 4. If a particular application does not require the full sample size, however, one could use only Wave 1 data for October and multiply weighted results by a factor of 4/3 to compensate for having only three-fourths of the sample. This is illustrated in figure 5.8.

To use Wave 1 data for the month of November, double the estimates (which compensates for having only one-half of the sample, consisting of Rotation Groups 3 and 4); for December, multiply the estimates by 4 (since they are based on a one-fourth sample consisting of Rotation Group 4 alone). Corresponding factors apply to data for June, July, and August (also available in Wave 1) as well; for these months the factors must be used, as the alternative of picking up the missing rotation groups in another wave does not exist.

A similar approach is applicable to subsequent waves as well. The particular factor to use is determined by the number of rotation groups covered in the time period one is analyzing. Factors for Waves 1 and 2 and combined Wave 1 and 2 estimates are given in figure 5.9.

Caveats for Estimation

SIPP cross-sectional files support monthly cross-sectional analysis. However, there are two caveats the user should keep in mind.

FIGURE 5.8 - FACTORS FOR MONTHLY DATA: WAVE 1, 1984 PANEL

Month of Interview	Rotation Group	Reference Period: Second Quarter Apr.	May	June	Third Quarter July	Aug.	Sept.	Fourth Quarter Oct.	Nov.	Dec.
October	1			X	X	X	X			
November	2				X	X	X	X		
December	3					X	X	X	X	
January	4						X	X	X	X
Factors to Compensate for Missing Rotation Groups				4	2	4/3	1	4/3	2	4

First, evaluations of responses to income and related variables that are provided on a monthly basis indicate there may be some biases in this reporting. For example, people tend to report a change in recipiency more often between waves than within waves (Burkhead and Coder (1985), and Weidman (1986)). This suggests it may not be possible to pinpoint recipiency changes to a specific month.

Second, most data users have been able to work only with annual income figures to this point, using the census, CPS, or other surveys which measure income only once during a year. There will be considerable temptation for SIPP users to return to familiar analytical ground by using household composition at the time of the interview and multiplying 4-month income figures by 3 to estimate 12-month income. To do so would ignore seasonal variation in employment and income, and observed changes in unit composition. A better approach to annual income would be to match together the three rounds of interviewing that cover a calendar year, and look at actual income experience across 12 months, perhaps comparing the results to the annual income and taxation information collected twice in each panel.[7] Note, however, that construction of annual income for families and households requires the construction of longitudinal units.

FIGURE 5.9 - FACTORS TO BE APPLIED TO MONTHLY DATA TO OBTAIN MONTHLY ESTIMATES

Estimates from Wave 1 Public-Use File	
June 1983 and December 1983	4.00
July 1983 and November 1983	2.00
August 1983 and October 1983	1.33
September 1983	1.00
Estimates from Wave 2 Public-Use File	
October 1983 and March 1984	4.00
November 1983 and February 1984	2.00
December 1983 and January 1984	1.33
Estimates from Combined Wave 1 and Wave 2 Public-Use Files	
June 1983 and March 1984	4.00
July 1983 and February 1984	2.00
August 1983 and January 1984	1.33
September through December 1983	1.00

[7] Linking of public-use products is described in chapter 6.

Time-Dimensioned Summary Statistics

An approach to analyzing these data that would reduce the biases for monthly estimates (discussed above) involves summarizing data across time. In this approach, one calculates standard summary statistics such as counts, means, and modes across time, as well as across individuals.

For example, instead of calculating the number of persons with incomes over $3,000 for the month of July, one would calculate the number of persons with a mean monthly income of $3,000 or more during the third quarter.

This approach is relatively straightforward at the person level. However, at the family or household level, an additional complexity is added because one must first define these groups longitudinally and devise a method to deal with units that may not exist for the full period.[8]

PREPARING ESTIMATES FOR DIFFERENT GEOGRAPHIC UNITS

Estimates for Census Regions

The total estimate for a region is the sum of the state estimates in that region. However, one of the groups of states, formed for confidentiality reasons, crosses regional boundaries. This group consists of South Dakota (Midwest Region), Idaho (West Region), New Mexico (West Region), and Wyoming (West Region). To compute the total estimate for the Midwest Region, a factor of .203 should be applied to the above group's total estimate and added to the sum of the other state estimates in the Midwest Region. For the West Region, a factor of .797 should be applied to the above group's total estimate and added to the sum of the other states in the West.

Estimates for regions included in the published SIPP reports reflect the actual region of residence, not the results of proration across the four-state group. Thus, there will be minor discrepancies between published regional totals and estimates derivable from microdata files for the Midwest and West regions.

Estimates for States

Estimates from this sample for individual states are subject to very high variance and are not recommended. The state codes (see figure 4.7) on the file are primarily of use for linking respondent characteristics with appropriate contextual variables (e.g., state-specific welfare criteria) and for tabulating data by user-defined groupings of states.

Estimates for the Metropolitan Population

Because of disclosure risk, metropolitan location cannot be revealed for all persons with metro residence. For 15 states in the SIPP sample, metropolitan or nonmetropolitan residence is identified. These states are those listed in figure 5.10 with a factor of 1.0 for use in state or Metropolitan Statistical

---

[8]These problems do not arise in Wave 1, as households were defined as of the interview and changes during the reference months were not recorded. (See McMillen and Herriot, 1985.)

FIGURE 5.10 - METROPOLITAN SUBSAMPLE FACTORS: 1984 PANEL

(Multiply these factors times the weight for the person, family, or household)

		Factors for Use in State or MSA Tabulations of Metropolitan Population	Factors for Use in Regional or National Tabs of Metropolitan Population
Northeast:	Connecticut	1.0390	1.0432
	Maine	-	-
	Massachusetts	1.0000	1.0040
	New Jersey	1.0000	1.0040
	New York	1.0110	1.0150
	Pennsylvania	1.0025	1.0065
	Rhode Island	1.2549	1.2599
Midwest:	Illinois	1.0232	1.0310
	Indiana	1.0000	1.0076
	Iowa	-	-
	Kansas	1.6024	1.6146
	Michigan	1.0000	1.0076
	Minnesota	1.0000	1.0076
	Missouri	1.0611	1.0692
	Nebraska	1.7454	1.7587
	Ohio	1.0134	1.0211
	Wisconsin	1.0700	1.0782
South:	Alabama	1.1441	1.1511
	Arkansas	1.0000	1.0061
	Delaware	1.0000	1.0061
	District of Columbia	1.0000	1.0061
	Florida	1.0333	1.0396
	Georgia	1.0000	1.0061
	Kentucky	1.1124	1.1192
	Louisiana	1.1470	1.1540
	Maryland	1.0000	1.0061
	North Carolina	1.0000	1.0061
	Oklahoma	1.1146	1.1214
	South Carolina	1.1270	1.1339
	Tennessee	1.0000	1.0061
	Texas	1.0192	1.0254
	Virginia	1.0778	1.0844
	West Va. - Miss.	-	-
West:	Arizona	1.0870	1.0870
	California	1.0000	1.0000
	Colorado	1.0000	1.0000
	Hawaii	1.0000	1.0000
	Oregon	1.0779	1.0879
	Washington	1.0868	1.0868

- indicates no metropolitan subsample is shown for the states.

Area (MSA) tabulations. (Use of these factors is explained below.) (On the rectangular file, use variable H*-METRO to identify households in the metropolitan sample. On the complex file, use METRO.) Metropolitan residence is defined according to the definition of MSA's as of June 30, 1983. In 21 additional states, where the nonmetropolitan population in the sample was small enough to present a disclosure risk, a fraction of the metropolitan sample was recoded so as to be indistinguishable from nonmetropolitan cases (METRO=2). These are the remaining states in figure 5.10. In these states, therefore, the cases coded as metropolitan (METRO=1) represent only a subsample of that population. Producing nonmetropolitan estimates is not appropriate using the SIPP public-use files.

In producing state estimates for a metropolitan characteristic, multiply the individual, family, or household weights by the metropolitan inflation factor for that state, presented in figure 5.10. (This inflation factor compensates for the subsampling of the metropolitan population and, as noted above, is 1.0 for the states with complete identification of the metropolitan population.)

In producing regional or national estimates of the metropolitan population, it is also necessary to compensate for the fact that no metropolitan subsample is identified within two states (Maine and Iowa) and one state-group (Mississippi-West Virginia). (There were no metropolitan areas sampled in South Dakota-Idaho-New Mexico-Wyoming.) Therefore, a different factor for regional and national estimates is in the right-hand column of figure 5.10. The results of the regional and national tabulations of the metropolitan population will be biased slightly, although less than one-half of one percent of the metropolitan population is not represented.

Estimates for the metropolitan population produced from the microdata files will differ from the published summary figures, not only because of the subsampling scheme but also because of differences in the definition of the metropolitan population. Figures published in the reports for the third and fourth quarters of 1983 were based on Metropolitan Statistical Areas (MSA's) defined for the 1980 census. That definitional change resulted in increasing the metropolitan population by 1.4 percent. Starting with the first quarter 1984 report (P-70, No. 3), the published figures reflect 1983 MSA definitions, as do the microdata files.

Estimates for the Nonmetropolitan Population

State, regional, and national estimates of the nonmetropolitan population cannot be computed directly, except for the 15 states where the factor in figure 5.10 is 1.0. In all other states, the cases identified as not in the metropolitan subsample (METRO=2) are a mixture of nonmetropolitan and metropolitan households. Only an indirect method of estimation is available: first compute an estimate for the total population, then subtract the estimate for the metropolitan population.

REFERENCES

Burkhead, D. and J. Coder. "Gross Changes in Income Recipiency From the Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1985.

Carlson, S. and R. Dalyrmple. "Food Stamp Participation: A Comparison of SIPP With Administrative Records," SIPP Working Paper Series No. 8604. Washington, D.C.: U.S. Bureau of the Census, 1986.

Chapman, D., L. Bailey, and D. Kasprzyk. "Nonresponse Adjustment Procedures at the U.S. Bureau of the Census," Survey Methodology, Vol. 12, No. 2 (1986), pp. 161-180.

Duncan, G. and M. Hill. "Conceptions of Longitudinal Households--Fertile or Futile?" Journal of Economic and Social Measurement (formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 361-375.

Huang, H. "Obtaining a Cross-Sectional Estimate From a Longitudinal Survey: Experiences of the ISDP," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1984.

Kalton, G. and D. Kasprzyk. "The Treatment of Missing Survey Data," Survey Methodology, Vol. 12, No. 1 (1986), pp. 1-16.

McMillen, D. B. and R. A. Herriot. "Toward a Longitudinal Definition of Households," Journal of Economic and Social Measurement (formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), pp. 349-360.

Weidman, L. "Investigation of Gross Changes in Income Recipiency From the Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Associaton, 1986.

CHAPTER 6 - LINKING WAVES

This chapter provides basic guidelines on the procedures to be followed in linking waves of data from the Survey of Income and Program Participation (SIPP). SIPP is a longitudinal survey of persons; therefore, the link must necessarily take place at the person level. If longitudinal units such as families or program units are desired, they should be constructed as a secondary activity. However, the procedures for determining longitudinal units have not been clearly established due to the lack of consensus on the definition of what constitutes the same unit over time. Due to this lack of agreement, this chapter does not address the methods for constructing longitudinal units. For information regarding the issues involved, the reader is referred to McMillen and Herriot (1985) and Citro et al. (1986).

This chapter is subdivided into three sections; the first presents the conceptual issues associated with matching SIPP data across time. The second describes the steps involved in linking successive waves, including a description of a type of observation which requires special handling in this match. The final section describes the options for linking nonsuccessive waves of SIPP data.

This chapter assumes the user is starting with one of the two types of public-use cross-sectional files (see chapter 4 for a description) and that the user is sufficiently skilled in handling these files to produce a person-level extract containing information on the sample unit, the household, and the family in which each person resides in each month. Inexperienced users are referred to chapter 4 for the requisite information.

The discussion which follows presents the methodology to be employed with only occasional references to the appropriate software. A number of different existing software packages can be used to accomplish this merge. To date, we are aware that OSIRIS IV (distributed by the Institute for Social Research) and INGRES (a relational database management system) have been successfully used.[1] Another option is to develop the process using a procedural language such as FORTRAN or PL/1. To date, PL/1 has been successfully used for this purpose. If the user elects to conduct the match with an existing software package, care should be used in the development of selected phases of the project.

CONCEPTUAL ISSUES

While the mechanics of linking successive waves of cross-sectional files may be straightforward, there are a number of thorny conceptual issues to be considered in defining the universe for the merged file. These issues are also relevant from the standpoint of quality control in that they must be understood before the user can interpret statistics on nonmatches as correct.

---

[1] Information on the OSIRIS IV software package can be obtained from the Survey Research Center, University of Michigan, Ann Arbor, MI 48109; and information on the application of the INGRES database management system can be obtained from the SIPP ACCESS CENTER, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706.

The text which follows presents a discussion of the straightforward mechanics of merging files as well as the conceptual issues. Users should note that the most complex problems affect only a small proportion of the sample (as discussed below). Hence, considerable resources can be saved by ignoring these problems without major ramifications to the analysis.

Longitudinal Design

To recap chapter 2, the survey is intended to follow a sample of adults for over 2 1/2 years. This group of adults, referred to as original sample adults, is defined as all persons aged 15 and over in interviewed households in the first wave of each panel.[2] In order to enhance the knowledge gained about the characteristics of the original sample adults, the survey includes a provision to collect information on other adults and children with whom the original sample adults reside during the months they are followed. However, these additional persons are not necessarily followed over time, so complete information on their characteristics may not be collected.

There are exceptions to the extent to which original sample adults are followed. Clearly no information is collected beyond the date of death. However, the design of the survey will also prohibit the collection of information when these individuals move outside the scope of the survey. There are several relevant survey concepts to keep in mind in identifying individuals and in determining which individuals are followed and which are not. First is the concept of the sample unit. This is a grouping of original sample adults according to their residence at the time of the Wave 1 interview. The nine-character field "SU-ID" in the cross-sectional files identifies this unit. Children in those Wave 1 households are assigned the same value of SU-ID as the adults with whom they reside. New people who enter the survey after Wave 1 are assigned the value of SU-ID associated with the original sample adults with whom they reside when they enter the sample.

The second concept is entry address (PP-ENTRY). This is the physical unit the person resided in when they were first interviewed. Original sample adults and children residing with them in Wave 1 will have an entry address identifier of 11. New persons who enter the sample are assigned the identifier of the physical unit in which they resided when first interviewed. This may be 11 or greater than 11.[3]

The third concept is the person number (PP-PNUM). Persons (including children) in interviewed households in Wave 1 are assigned numbers sequentially, starting with 101. Persons who enter the survey after Wave 1 are assigned numbers in excess of 200. In each household in each wave these numbers are assigned to

---

[2]Exceptions to this rule are persons erroneously interviewed in Wave 1, persons in households scheduled for interview in Wave 1 but not successfully contacted until Wave 2, and children in Wave 1 households who attain age 15 before moving away from other Wave 1 adults.

[3]This is derived from the variable address identifier (ADDID) which is described in more detail in chapter 5.

new individuals sequentially starting with 1+(wave*100). For example, if two people enter a household in Wave 2, their person numbers would be assigned the values of 201 and 202.

The three variables described above (sample unit identifier, entry address identifier, and person number) are required in combination to uniquely identify each individual across waves in this survey. Hence, all of these variables must be used to link successive waves of data. Furthermore, the person number in combination with age is required to distinguish persons who are followed in the survey and those who are not. Children (who are not followed) are persons whose age at the time of the interview is less than 15.[4] For adults, the value of the person number can be used to distinguish between those who are followed (person number less than 200) and those who are not (person number greater than 200).

Note there are two exceptions to the rule for uniquely identifying individuals over time in SIPP. Occasionally original sample adults and, possibly, their children may move into a household with one or more original sample adults who have a different sample unit identifier. The newly formed household is called a merged household. In this case, the sample unit identifier and person number are changed for the sample unit movers. The sample unit identifier for one set of original sample persons is set equal to the sample unit identifier of the persons they joined, and person numbers are assigned sequentially starting with 180. This person number assignment is intentionally set to be less than 200 in order to identify these individuals as original sample members.

The second exception is that a new sample adult can have the value of the person number altered if a sample unit splits and a new person joins each of the succeeding units. The need to change person numbers arises when the sample unit adults reunite and both new adults remain in the reunited household. Left unaltered, the two new adults would have the same person number. Hence, to avoid confusion, one new adult has the value of the person number changed.

Consider, for example, an intact household headed by a married couple who separate in Wave 3, each moving in with a sibling. Both siblings are assigned a person number 301 because they enter the sample in Wave 3 at different addresses. If the husband and wife reunite in Wave 6 bringing the siblings with them, then two persons with the person number 301 would be residing in the reunited household unless one of the numbers is changed. Note that the new value of the person number in this case is in the range 680 to 689 because the person number was changed in Wave 6.

These cases are cumbersome to deal with in the link and are, therefore, discussed in a separate section of this chapter ("Special Treatment of Sample Unit Movers" in the "Procedure for Linking Successive Waves" section). The following subsections which describe the possible outcomes of a match between two successive waves include several specific references to this situation and how to cope with it.

---

[4]Children who reach age 15 during the life of the panel will be followed subsequent to their 15th birthday regardless of whether they continue to reside with an original sample adult.

Within-Wave Classification

As a by-product of the design of the SIPP survey and data products, there are a number of different types of outcomes which can occur when linking successive waves. In order to understand these, it is important first to recognize the various situations which can exist within one wave because the outcomes across waves are essentially transitions among the various within-wave situations. Chapter 5 outlines the various ways in which individuals can be classified within a wave according to their attachment to the sample and the nature of the outcome of attempts to interview them. The user should study that chapter before proceeding with the rest of this chapter.

Cross-Wave Transitions

The possible situations which can arise when linking two successive waves, N and N+1, are described in figure 6.1 using the categorization of within-wave situations defined in figure 5.2 of chapter 5. The majority of cases are original sample members interviewed in both waves (AA) and, hence, characteristics of those individuals over time can be analyzed from the reported data. Care should be taken in this analysis, however, because of two factors. First, item nonresponse can exist on one or both of the matched interview records and the missing values will have been imputed cross-sectionally (see chapter 7). Second, basic demographic information such as age, race, and sex can be corrected subsequent to Wave 1, making the data inconsistent over time. The value appearing in the most recent wave is believed to be correct.

Original sample members interviewed in Wave N may change sample units before the Wave N+1 interview (AC). These cases are distinguished from the first group by the fact that the basic person identifiers change and there may be two records for each of the individuals in Wave N+1. These two factors increase the difficulty of the link. Therefore, the procedure for handling these situations is described in "Special Treatment of Sample Unit Movers" in the "Procedure for Linking Successive Waves" section. Once the data are properly linked, however these individuals can be treated in the same manner as the first Group (AA) with the same precautions regarding cross-sectional imputations for item nonresponse.

The next five transitions, listed in figure 6.1, describe the various ways in which a noninterview for original sample adults interviewed in Wave N can be treated in Wave N+1. Groups AD, AE, and AF represent situations where the economic information in Wave N+1 has been imputed cross-sectionally so that changes in characteristics over time for these records are not real. These cases should be classified separately or excluded in a longitudinal analysis of individuals. Group AD represents a situation where the individual is considered "in sample" for all months in the two waves. Groups AE and AF represent situations where the individual is only considered "in sample" for part of the Wave N+1 reference period, with the distinction between them being that interviews with Group AF people are not attempted in Wave N+1 because they are no longer in the sample. Group AE represents unsuccessful attempts to interview individuals who are expected to have remained within the scope of the survey. In both groups (AE and AF), the economic data will be imputed for the entire Wave N+1 record, but information for months "not in sample" should be ignored, even though they are present.

FIGURE 6.1 - POSSIBLE TRANSITIONS BETWEEN TWO CONSECUTIVE WAVES, N AND N+1 (Based on within-wave categories from chapter 5)

WAVE N	WAVE N+1	DESCRIPTION
A	A	Original sample adults interviewed in both waves
A	C	Original sample adult moves to a new sample unit in Wave N+1 and changes identifiers (there may be two records in Wave N+1 for this person[1])
A	D	Original sample adult refuses to respond in Wave N+1 (the entire questionnaire for Wave N+1 is imputed)
A	E	Original sample adult moves and is not successfully interviewed in Wave N+1 but is classified as "in sample" at least one month in the later wave (the entire questionnaire for Wave N+1 is imputed)
A	F	Original sample member leaves the sample after the middle of the first reference month of Wave N+1 (the entire questionnaire for Wave N+1 is imputed)
A	G	Original sample member who along with all members of the entire Wave N and Wave N+1 households was not interviewed in Wave N+1
A	N (No record)	Original sample member who left the sample before the middle of the first month of the Wave N+1 reference period or who was excluded due to a deliberate modification in the sample[2]
B	B	New sample member remains in the sample and is successfully interviewed in both waves
B	C	New sample member remains in the sample but moves to a new sample unit, or remains in a reunited unit where another person exists with the same person number (identifiers are changed in Wave N+1 and there may be two records for this person[1])
B	D	New sample member remains in the sample but refuses to be interviewed (entire questionnaire is imputed in Wave N+1)
B	F	New sample member leaves the sample after the first half of the first reference month of Wave N+1 (entire questionnaire is imputed)
B	N (no record)	New sample member is excluded from Wave N+1 due to a deliberate modification in the sample or because he/she left the survey before the first half of the first reference month of Wave N+1

FIGURE 6.1 - CONTINUED

C	A or B	Sample unit mover who changed sample units in Wave N continues to be interviewed in Wave N+1 under the new identifier[3]
C	D	Sample unit mover becomes a Type Z nonrespondent (Wave N+1 questionnaire is imputed)
C	E	Sample unit mover moves again and is not successfully interviewed in Wave N+1 (Wave N+1 questionnaire is imputed)
C	F	Sample unit mover leaves the sample in Wave N+1 (Wave N+1 questionnaire is imputed)
C	G	Sample unit mover along with all persons in both the Wave N and N+1 households was not interviewed in Wave N+1
C	N (no record)	Sample unit mover in Wave N who left the sample before the middle of the first reference month of Wave N+1 or who was excluded because of a deliberate modification in the sample
D	A or C	Type Z nonrespondent in Wave N is successfully converted to a response in Wave N+1 (Wave N questionnaire is imputed)
D	D	Type Z nonrespondent continues to refuse to be interviewed in Wave N+1 (Wave N and N+1 questionnaires are imputed)
D	F	Type Z nonrespondent leaves the sample (questionnaires are imputed)
D	G	Type Z nonrespondent in Wave N is not interviewed in Wave N+1 along with an entire household (data are imputed in Wave N and missing in Wave N+1)
D	N (no record)	Individual is excluded from Wave N+1 due to a deliberate modification in the sample or because he/she left the sample before the middle of the first month of the reference period
E	A	An unlocated mover in Wave N was successfully interviewed in Wave N+1
E	N (no record)	An unlocated mover in Wave N was dropped from the sample when not found in Wave N+1 or due to a deliberate modification in the sample
F	N (no record)	Person left the sample in Wave N and was not followed[4]

FIGURE 6.1 - CONTINUED

G	A or B	At least one member of a Type A noninterviewed household in Wave N is successfully interviewed in Wave N+1 and an interview was obtained for this record in Wave N+1
G	D	At least one member of a Type A noninterviewed household is successfully interviewed in Wave N+1 and this person was a Type Z nonresponse in Wave N+1
G	N (no record)	A Type C noninterview household in Wave N, or the person was excluded from Wave N+1 due to a deliberate modification in the sample. Also occurs when no member of a Type A noninterviewed household was successfully interviewed in Wave N+1.
H	A	Child was included in the sample in Waves N and N+1 and reaches the age of 15 before the Wave N+1 interview
H	H	Child was successfully identified as present with an original sample adult in both Waves N and N+1
H	E, G	Child was successfully identified as present in Wave N and at least one month of Wave N+1 but resided in a household in Wave N+1 in which no interview was successfully conducted
H	N (no record)	Child was not present with an original sample adult in Wave N+1
N (no record)	A	Original sample member with no record in Wave N due to a deliberate modification in the sample
N (no record)	B	New sample member enters in Wave N+1 (this may only ever occur in the first month of the reference period)[5] or has no record in Wave N as the result of a deliberate modification in the sample
N (no record)	H	Child moves into (or is born into) an original adult sample member's household or has no record in Wave N as the result of a deliberate modification in the sample

[1] There are two records if the move took place after the middle of the first month of the reference period. One record has the old ID (and the complete questionnaire is imputed) and the other record has the new ID with reported data. Otherwise, there is one record with the new ID.

[2] There is also a reduction in sample between Waves 1 and 2 in each panel due to the elimination of adults who were incorrectly interviewed in Wave 1. A list of these is provided in the User Note Series available from Customer Services, Data User Services Division, Bureau of the Census, Washington, D.C. 20233 (301-763-4100).

[3] In this case only one record exists in Wave N+1 and it has the new identifier.

[4] Some of these individuals may return to the sample after they leave.

[5] In some cases the new sample member was actually present during the interview in Wave N but no Wave N record was constructed because it was determined that the person entered the sample household after the Wave N reference period. The person numbers for these exceptional cases fall between N*100 and (N+1)* 100.

The transition from Category A to Category G represents a situation where record exists in Wave N+1 for an original sample adult interviewed in Wave for whom virtually no information on the characteristics exists in the Wave N reference period. These individuals are members of Type C noninterviewed hous holds or they, along with all other persons they lived with in both the Wave and the Wave N+1 reference periods, were in a Type A noninterview household Wave N+1. Records for these individuals should either be classified separate or ignored altogether in a longitudinal study. (See chapter 7 for the definiti of types of interview.)

The final transition possible for original sample adults interviewed in Wave is AN, where no record exists in Wave N+1. In most cases this is due to deliberate modification in the sample between the two waves. This modificati could either be a permanent reduction in the sample size in order to conser costs (as was the case between Waves 4 and 5 in the 1984 Panel) or a simp shift in the interviewing cycle. The latter refers to the reduction of select waves of interviewing to three rotation groups (such as occurred in Waves and 8 of the 1984 Panel). This does not mean there is a gap in the data f the excluded group, but rather that the continuing information was collected Wave N+2. Other cases of a Type AN transition are people mistakenly interview in Wave 1 and later dropped from the sample.

The recommended approach for the observations affected by a deliberate modif cation in the sample between Waves N and N+1 or the mistaken Wave 1 intervie is to eliminate these cases. The exception to the rule is when a shift interviewing cycle occurred. In that case, information for rotation grou excluded from Wave N+1 could be appended to the linked file by merging t data from Waves N and N+2. In this case, Wave N+2 can be treated in the sa manner as Wave N+1.

The remaining situation which can cause an original sample adult interview in Wave N to have no record in Wave N+1 is that the individual left the samp in the first half of the first reference month in Wave N+1. If the longitudin analysis includes observations on those individuals who left the sample, the records should be kept in the linked file; otherwise, they should be ignore

The next five groups of transitions in figure 6.1 (BB, BC, BD, BF, BN) descri the possible situations which can occur with adults who are not part of t original sample and, hence, are not followed when they no longer reside with original sample adult. The general issue to consider is whether to inclu these individuals at all in a longitudinal person-based study. If attribut of families and households are desired, information on these individua should not be discarded. However, it may not be desirable to consider the individuals as separate analysis units because of the absence of informati on their characteristics for the waves when they did not reside with an origin sample adult. To the extent that these new sample members are included in longitudinal study in order to construct monthly attributes such as household and family income, the user should be aware of when the data are imputed a when they are not. This is described in figure 6.1.

The next six transitions listed in figure 6.1 (CA, CD, CE, CF, CG and C describe what can happen to individuals who change sample units in Wave Note that because there can be two records for each of these individuals Wave N, two different transitions can be observed for essentially the sa

person. The important fact to remember is that the continuing record in Wave N+1 will use the new person identifiers in the main identification fields rather than the old (the old are stored elsewhere in the record). Hence, a match will occur under the new identifier but not the old. Also note that the continuing record for these individuals in earlier waves will use the old identifier. As mentioned earlier, the alternative approaches for conducting the link for sample unit movers is described in a separate section. However, once the data are appropriately merged, the issues involved in the analysis of these individuals are the same as those described earlier for original sample adults, with the same precautions regarding imputed data.

Transitions DA through DN in figure 6.1 describe the possible outcomes for Type Z nonrespondents (original sample adults in a household where other adults responded to the survey) in Wave N. Essentially they can be converted to an interview in Wave N+1 (DA or DC) with some possibility of item nonresponse, which is compensated for cross-sectionally. They can also continue to refuse to participate (DD or DG), or they can leave the sample altogether (DF or DN). In all of these last four situations, Wave N data are imputed. In two of the four situations (DD and DF), the Wave N+1 data are also imputed. Figure 6.2 displays the outcome for Type Z nonresponse in the first three waves of the 1984 Panel. For longitudinal analysis, it is recommended that these individuals either be excluded or treated as a separate category. This is due to the fact that the number of cross-sectional imputations obscures any real transitions which may occur across waves of interviewing.

Persons who were unsuccessfully interviewed in Wave N (Category E) can only have two possible outcomes in Wave N+1. Either they can be interviewed (EA), or they will be dropped for the sample (EN). As is true for all cases where some information is missing (and possibly cross-sectionally imputed), these cases should either be eliminated from the study or treated separately.

Some individuals will have naturally left the sample in Wave N (Category F). There should be no record for them in Wave N+1 and a nonmatch is valid (Type FN transition). Treatment of these individuals is a function of the scope of the longitudinal analysis. This group includes persons who died (and hence should be kept in the study), as well as those who no longer remain within the scope of the survey, such as those who are institutionalized.

Children have four possible transitions between waves. They can be in the sample in both waves and reach the age of 15 before the Wave N+1 interview, in which case they are treated as original sample adults (HA). They can continue to reside with an interviewed sample adult in both waves; hence, there are records in both waves which will match (HH).[5] They can reside with an interviewed adult for the early part of the Wave N+1 reference period, in which case a record exists in both waves with "in sample" codes set for the time they were with the interviewed adult (HE or HG). Alternatively, there will be no record in Wave N+1 (HN). It is recommended that children be treated in the same manner as other sample members who are not followed.

[5] In the event that a child with a 200+ person number moves with an adult, that child's person number may change.

FIGURE 6.2 - UNWEIGHTED COUNTS OF TYPE Z NONRESPONDENTS* ACROSS WAVES 1, 2, 3 OF THE 1984 PANEL, ROTATION GROUPS 1, 2, and 3

WAVE	NOT CONVERTED 1/	CONVERTED IN WAVE N+1 2/	CONVERTED IN WAVE N+2 2/
1	165	120 3/	23
2 4/	216	78	NA
3 5/	365	NA	NA

*Original sample adults in a household where other adults responded to th survey.

1/ Includes those who leave the sample as well as those who remained Type Z
2/ Converted to either self or proxy interview.
3/ Includes 32 persons who left the sample in Wave 3 or became a Type nonrespondent.
4/ Includes 112 people who were Type Z in Wave 1.
5/ Includes 130 people who were Type Z in Waves 1 or 2.

The final transitions described in figure 6.1 (NA, NB, NG) describe situation which represent valid nonmatches of records for individuals in Wave N+1. On cause of this occurrence is the modification in the interviewing cycle fo selected waves, resulting in the fact that one rotation group is excluded fro Wave N. If the user decides to retain the affected cases, relevant informatio can be extracted from Wave N-1.

Other causes of this type of nonmatch are the entry of adults or children int an original sample member's household, or the move of an original sample membe into a nonsample member's household. The recommended treatment of these ne sample members is to include information on them in constructing attributes o sample adults' households and families, but not to treat them as separat analysis units.

Unit of Analysis

As stated in the introduction, the SIPP survey is a longitudinal survey o adults interviewed in Wave 1 (original sample adults). Although data fo children and other adults with whom they reside are collected, the design doe not guarantee full information on their characteristics over an extended perio of time, as illustrated in the preceding section. Furthermore, due to sampl attrition, noninterview, and movement, incomplete data for original sampl adults may occur. These situations create difficulties in the analysis o linked data because of incomplete observations. For discussions of alternativ methods of treatment of incomplete records, the user is referred to Kalton et al. (1987). One straightforward approach is to choose "original sampl adult" as the principal unit of analysis. Children and other adults coul be retained in the file but used primarily for the calculation of household and family characteristics.

Potential Problems in Cross-Wave Linked Files

While the survey is truly longitudinal, the public-use cross-sectional products have not been intended for use as longitudinal data. Hence, a number of potential problems exist in the development and analysis of cross-wave linked files. These are discussed below.

1. Weights

The sample design is complex so analysis should not be performed using unweighted observations. The weights available for use on the cross-sectional files can be used for this purpose although they are not true longitudinal weights, and the choice of an appropriate one varies depending on the unit of analysis. The cross-sectional files contain five monthly weight fields for each person, one for the interview month and one for each reference month. These are nonzero in each month the person is considered "in sample." These weights have been adjusted so that aggregate population statistics for each month agree with independent estimates for the total population (see chapter 5 for a discussion of their derivation). In Wave 1 there is also a weight for the household (called HH-BW) which is essentially the inverse of the Wave 1 sampling ratio for each household adjusted for noninterviews in Wave 1.

The user can use the initial Wave 1 weights if the longitudinal analysis is solely based on the members of interviewed Wave 1 households. However, sample attrition occurs and it is not random so these weights do not properly account for this form of nonresponse. Another alternative is to request calendar-year weights developed by the Census Bureau.[6] In the event that the calendar-year weights are selected, the user should restrict the sample for longitudinal analysis to those persons included in the calendar-year weighting process.[7]

2. Imputations

To reiterate the precautions noted in "Cross-Wave Transitions," the user should remember that the data products are cross-sectional. This means that, with the exception of the items derived from the control card, all edits and imputations were performed on each wave as if there were no other source of information about the sample members. Hence, caution should be used when analyzing persons for whom any items were imputed at any point in time. In order to completely determine whether any items from the questionnaire have been imputed, the user should refer to the following variables in each wave:

PP-MIS	If PP-MIS-5=2 ("not in sample") and PP-MIS-M=1 for any month (M=1, 2, 3, 4), then all of the questionnaire items were imputed

---

[6]Requests for the "1984 Calendar-Year Annual Weight File" should be sent to Customer Services, Data User Services Division, Bureau of the Census, Washington, D.C. 20233 (301-763-4100).

[7]Memorandum dated 4/29/87 from Gary Shapiro to Thomas Walsh on the subject "SIPP 1984-Specifications for CBO Weighting."

PP-INTVW	Values of 3, 4 denote Type Z noninterview (Category D) where all of the questionnaire items were imputed
IMPUTATION FLAGS	These denote item imputation

Demographic data derived from the control card may be inconsistent acros waves even though these data were processed longitudinally. This occurs whe the interviewer believes the control card information is incorrect and change the data. It is recommended that the most recent control card information b used when there is an inconsistency across waves.

3. Interview Month

Selected information pertaining to the interview month of Wave N is collecte twice, once in Wave N and again in Wave N+1. Inconsistent information on house hold composition and selected demographic data can exist either because of th timing of the interview or because of the correction of control card data i Wave N+1. The user is advised to use the data collected in Wave N+1.

PROCEDURE FOR LINKING SUCCESSIVE WAVES

The recommended procedure for linking two successive waves, N and N+1, consist of five separate steps. The first two steps are the preparation of extract from the full cross-sectional files and the execution of two sorts. Next the data are linked, followed by the verification of the merge. The fina step (which is actually a series of activities which can be carried out a different points in the process) is to deal with sample unit movers. Each o these is discussed below. The recommended procedure for linking three or mor successive waves (N, N+1..., N+K) is to first link Waves N and N+1, then lin the output of that to Wave N+2, and continue in that manner until all wave have been added.

Initial File Preparation

To link Waves N and N+1 first, prepare extracts from the respective waves These extracts should be person-based files containing, at a minimum, the vari ables from figure 6.3 which are needed to uniquely identify each person and t determine the classification of each record within the wave as described i "Longitudinal Design" in the "Conceptual Issues" section. Additional variable to be added are a function of the analysis to be performed. For example, i the linked file is to be used to analyze changes in program participation ove time, the extracts should contain the relevant monthly program participatio flags and benefit amounts, as well as the variables to be used as stratifier (in a descriptive analysis) or independent variables (in a regression model)

On the other hand, if the ultimate goal is to produce a general purpose linke file, the recommended approach is to prepare a small extract of the two wave containing the variables in figure 6.3, plus the following:

SUSEQNUM (Sequence number of the sample unit within each wave
PP-RCSEQ (Sequence number of each person within sample unit c each wave)

The link can then be constructed cheaply using the small files, and the remainder of the data elements can be added by merging other data of interest from the cross-sectional files to the linked file. The latter should be carried out using SUSEQNUM and PP-RCSEQ.

In the linking of two successive waves, there are several possible outcomes for each individual surveyed:

1) There is a record in both waves and they match.

2) There is a record in Wave N with no corresponding match in Wave N+1.

3) There is a record in Wave N+1 but no match in Wave N.

4) An individual changes sample units.

Depending on the analysis to be performed, selected individuals in groups 1, 2, or 3 may be ignored. To the extent it is feasible, costs can be reduced by excluding these individuals from the extract. A description of the types of cases which the user may want to exclude from the analysis, along with methods for identifying them at the extract step, is contained in figure 6.4. Cases which fall into group 4 need special handling in the link. Although some of the special work can be performed at the extract phase, the treatment of these individuals is discussed in a separate section below.

As discussed in "Merge Step" below there is a potential to create duplicate variable names on the matched file output, depending on the file structure chosen for the merged file. If this is going to be a problem, then variable names could be changed when these initial extracts are created in this first step.

Sort Step

Each of the extracts should be sorted prior to the actual link. The sort order is as follows:

SU-ID
PP-ENTRY
PP-PNUM

The second two fields can either be treated as two variables or concatenated as one five-character field. The latter is recommended. Sort utilities such as DFSORT and SINCSORT, available on IBM mainframes, can be used for this step.

Merge Step

The sorted extracts should be linked using the three sort-key fields noted above. In this step, an allowance should be made for the nonmatches which will occur whenever group 2 or group 3 cases remain in the extracts. The output of the merge step should be a microdata file containing the data elements from both the Wave N and N+1 extracts. There are several options for the structure of this file, the choice of which depends on the type of the analysis to be performed on the output and on the volume of data to eventually be included

FIGURE 6.3 - VARIABLES NEEDED TO IDENTIFY INDIVIDUALS AND TO DETERMINE THE CLASSIFICATION OF EACH RECORD

RECTANGULAR FILE	COMPLEX FILE	DESCRIPTION
SU-ID	SU-ID	Sample unit identifier
ITEM36B[1]	ITEM36-B[2]	Household interview status (values of 2-6,23 signify Category G, Type A and C noninterview, and values of 24, 25 signify Category E, noninterviewed movers)
PP-INTVW	PP-INTVW	Person interview status (values of 3 or 4 denote Category D, Type Z nonresponse)
PP-MIS1 PP-MIS2 PP-MIS3 PP-MIS4 PP-MIS5	PP-MIS1 PP-MIS2 PP-MIS3 PP-MIS4 PP-MIS5	"In sample" indicators for each month in the wave (1 denotes "in sample," 2 denotes "not in sample")-- Persons with PP-MIS5=2 and PP-MISJ=1 for at least one of the first 4 months have all questionnaire information imputer
PP-ENTRY	PP-ENTRY	Entry address ID
PP-PNUM	PP-PNUM	Person number (values greater than 200 denote Category B, new sample members who are not followed over time)
AGE-5	AGE-5	Edited age as of the interview date (persons with AGE-5 < 15 in Wave N who may not appear in Wave N+1 -- Category H)
U-REALFT[3] U-REAENT	U-ENTLFT	Reason for changing households (values of 5, 6, 7, 8 and 10 denote Category F, i.e., persons who leave the sample by design)
PREV-ID PREV-ROT SC0064 SC0066	PREV-ID PREV-ROT SC0064 SC0066	Original sample unit ID, rotation group, entry address, and person number for sample unit movers (Category C)

NOTE: Persons in Category A have PP-PNUM < 200, PP-ENTRY = 11, and are not classified as Category B-H in the above description.

1/On the rectangular file there are four variables, N1ITM36B, N2ITM36B, N3ITM36B, and N4ITM36B. Any one will suffice for each different address the person had. The best one to use is the one for the most recent month. However, do not extract this variable for months when ADDID=0.

2/On the relational file there is a potentially different value of ITEM36B on each household record associated with each individual. The best one to use for this purpose is the one on the household record corresponding to the most recent address.

3/These variables are inappropriately documented on the rectangular files. Refer questions on their use to the Census Bureau.

FIGURE 6.4 - CASES WHICH MAY BE EXCLUDED FROM LONGITUDINAL ANALYSES

TRANSITION FROM FIGURE 6.3	METHOD OF IDENTIFYING
AD, AE, AF, AG	In Wave N+1, Categories D, E, F, and G can be determined using the procedure described in figure 6.3. Hence the Wave N+1 extract can be reduced. However, the user cannot identify these cases directly from Wave N, so it is easier to include them in both extracts and delete them after the merge step described in "Merge Step" in the "Procedure for Linking Successive Waves" section. In fact, the latter approach is recommended to facilitate comparison with the quality control statistics provided by the Census Bureau in the user note series, issued as an addendum to the technical documentation.
AN, BN, DN, EN, FN, GN, HN	The User Note Series distributed by the Census Bureau will contain a list of people in Wave N who are not in Wave N+1. The distinction between the Wave N categories can be determined as described in figure 6.3. Note that, even if this general group is excluded from the matched files, it may be advisable to keep records for deceased individuals.
BB, BD, BF	New sample members can be distinguished by the value of person number (PP-PNUM > 200).
DA, DD, DF, DG	In Wave N, Category D can be determined as in figure 6.3, as can Categories D, F, and G in Wave N+1. However, it is recommended that these cases be included in the extracts, merged, and then deleted after the fact, if they are to be deleted at all. This will facilitate quality control using the match statistics provided in the user note series.
EA, GA, GB, GD	This cannot be fully determined until records are linked.
HH, HE, HG	Children are less than age 15 at the time of the interview. Hence, if children will be deleted entirely from the matched file, age and month of birth relative to the interview date can be used to reduce the number of records in the extracts.
NA, NB, NG	The User Note Series describes the cases in Wave N+1 who are not in Wave N.

in the merged file. In fact, it may be appropriate to produce two microdata files if conflicting needs exist for the analysis. Furthermore, depending on how the extracts were initially created and the file structure chosen, there can be duplicate variable names on the merged file which would be likely to cause a problem for the analysis.

1. File Structure

The microdata file can either be a flat file with one record per person, a flat file with one record per person wave, a flat file with one record per person month, or a hierarchical file with one record per person containing all the constant information such as sex followed by a series of person-month records containing information which varies over time. The file can either be variable length or fixed length depending on the system requirements and on the restrictions imposed by the software which will ultimately be used to process the data.

The choice of the appropriate file structure depends on whether the analyst wants to study gross flows (e.g., the research may be targeted to the number of people who experience a change in program participation status) or net flows (e.g., the number of program participants in time t compared to the number of participants in time t + m). Analysis of gross flows and related issues, such as characteristics of spells of program participation or poverty, are most easily carried out on a file which contains one record per person. However, analysis of net flows and related issues are most easily carried out with information stored in person-month form.

Another issue to consider in choosing the appropriate file structure is the record length of the analysis file and the restrictions the system or software places on that record length. If a series of successive waves are linked or if a large number of data elements are extracted from each wave, a flat person file may become too large (in terms of number of characters per record), thus exceeding specified limits of the system or of the software needed to analyze the data.

The final issue to note is the impact of part-period people on the choice of the file structure. Part-period people are those determined to be "not in sample" for at least one month of the combined reference periods of all waves being linked. These include cases where whole records are absent in at least one of the waves, as well as those considered "in sample" for only part of one or more waves.

If these people are included on the output file and a flat person file is the desired file structure, then the user should consider whether fixed length records are desired or required. If they are, then missing data or blanks should be explicitly generated when one or more months of data are missing. If variable length records are permitted, then the user may want to suppress data for months "not in sample" in order to conserve on storage costs.

If part-period people are included in the output file and a person-month format is chosen, then the file size can be minimized by deleting information for months "not in sample." However, this results in a variable number of records for each person which may not be desirable. An alternative is to generate a record for all months of the combined reference period, generating missing data

for persons who do not have a record in one or more of the cross-sectional files.

2. Variable Names

If the file structure chosen is a flat person file, duplicate variable names can result for both monthly data and wave-specific data. If the user cannot cope with duplicate variable names, then the names of the data elements extracted from the cross-sectional files should be changed. This can either be done in step 1 when the extracts are created or in the merge step before the file is written out. If the software used to create the file relies on variable numbers or file locations, this renaming is essentially done automatically. However, the analyst may get confused if the associated data dictionary contains the same variable name for two different locations on the merged file.

Duplicate variable names arise because the cross-sectional files use the same naming convention both to denote substantive meaning and to denote the time dimension. For example, PP-MIS-1 means "month-in-sample" in the first month of the reference period on both Wave N and N+1. If no change is made on the output file, there will be two variables called PP-MIS-1. One refers to the first month in the Wave N reference period, and the other refers to the first reference month of the Wave N+1 reference period.

Duplicate names such as these can be avoided by changing the subscripts on monthly variables extracted from the two waves and adding subscripts to the wave-specific variables. There are several options to rename monthly variables, three of which are described here. First, they can be numbered sequentially beginning with the first reference month of Wave 1 or the first reference month of Wave N. (Wave 1 is recommended as the starting point--particularly if Wave 1 is included in the link or if the user is uncertain how many or which waves will ultimately be included in the merged file.) The second choice is to use a two-digit subscript, the first of which denotes the wave and the second, month within wave on the cross-sectional files. A third choice, which is only recommended if calendar-month statistics are produced exclusively, is to use a three-digit subscript, the first digit of which denotes the year, and the second and third, the calendar month. Note that the choice of the third option implies a different set of subscripts for each rotation group within a wave because of the use of staggered interviewing.

For the wave-specific information, the user can elect to keep only one set (this applies to constant information such as sex) or to change the names in any convenient fashion such as adding the wave number as a subscript. Note, however, that even the constant information may change over time, so it may be advisable to retain a value from each wave with the name appropriately subscripted.

Verification

A table of the number of matches between successive waves and the number of nonmatched cases by reason for the nonmatch is provided by the Census Bureau in the User Note Series, distributed to persons who purchase the technical documentation. These statistics should be used for comparison purposes to ensure

the match was successfully carried out. In order to carry out this verification process, the user should prepare the following statistics:

1) Number of matched records
2) Number of sample unit movers with a record in Wave N+1 but no corresponding record in Wave N
3) Number of sample unit movers with a record in Wave N but no corresponding record in Wave N+1
4) Number of other people in Wave N with no corresponding match in Wave N+1, stratified by rotation group
5) Number of other people in Wave N+1 with no corresponding match in Wave N, stratified by rotation group and person number

Note that determining the sample unit movers is most easily accomplished by checking the person identifiers against the list of sample unit movers provided in the User Note Series.

A summary of the verification statistics for Waves 1 though 4 of the 1984 Panel is included in figure 6.5. For more specific information on sample unit movers in those waves, the user is referred to the User Note Series in Waves 2, 3, and and 4.

Special Treatment of Sample Unit Movers

On rare occasions, persons initially interviewed in one sampling unit will move in with individuals originally interviewed in another sampling unit. For example, in the 1984 Panel 14 persons were involved in such merges. When this happens, the identification variables for the sample unit mover are modified.

FIGURE 6.5 - VERIFICATION STATISTICS FOR MATCHING SUCCESSIVE WAVES IN THE 1984 PANEL

		IN WAVE N NOT IN WAVE N+1		IN WAVE N+1 NOT IN WAVE I		
N, N+1 (Wave)	In Both, Wave N And Wave N+1	Eliminated Rotation Group	Other	Not New In Wave N+1[1]	New Sample Member	Eliminate Rotatio Group
1-2	39891	13706	129	1	1551	N/A
1-3 (rotation group 4 only)	13652	N/A	54	3	517	N/A
2-3	40562	N/A	881	19	1444	14172
3-4	53922	N/A	2275	70	2001	N/A

[1]/This includes all persons whose person number is less than (N+1)* 100

The sample unit identifier (SU-ID) is changed to equal the sample unit identifier of the new merged household. So that these sample unit movers can still be identified as original sample members and at the same time retain unique identifiers, the value of the person number (PP-PNUM) is set to exceed 179 but not be greater than 199. The entry address identifier (PP-ENTRY) continues to be 11. Hence, in order to construct a longitudinal record for this type of mover from the cross-sectional files, the values of SU-ID, PP-ENTRY, and PP-PNUM must be explicitly modified. Furthermore, care must be taken when the mover moved in the middle of the reference period. In that case, there will be two records for that person in the wave in which the move occurred, one with the old sample unit identifier and one with the new. Note that this can occur in either Wave N or N+1. The record with the new identifier contains the responses from the interview, whereas the record with the old identifier contains imputed information. On the other hand, the demographic information from the control card and the weights are appropriately assigned on both records according to the date of the move. That is, the record with the old identifiers has nonzero data and is considered "in sample" prior to the move and the other record has nonzero data and is considered "in sample" subsequent to the move. In both cases, zeros are inserted in the demographic and weight fields for months considered "not in sample."

In the User Note Series, the Census Bureau provides a listing of the identifiers for persons who change sample unit, along with both the old and new identifiers. In the waves after the change, the microdata products will contain the original identifiers in the following fields.

PREV-ID (old SU-ID)
SCO064 (old PP-ENTRY)
SCO066 (old PP-PNUM

The new identifiers appear in the basic identifier fields as described above. There are several options to deal with the sample unit movers. One is to change the values

SU-ID
PP-ENTRY
PP-PNUM

on the wave(s) before the move occurred to the newly created identifiers. The other is to change the corresponding fields in the later wave(s) to match the original identifiers. The third option is to do both. The fourth option is to delete them from the analysis. There is a fifth option which is to do nothing, but that is not recommended.

The choice of one of these options is a function of the analysis to be performed. If the longitudinal analysis is strictly based on the individual, then either option 1 or option 2 is sufficient--the selection of which is a function of when the move occurred relative to the waves being linked. For example, if the move occurred between Waves 1 and 2 and there is a need to link more than two waves, then it would be more cost efficient to change the Wave 1 identifiers to match the later waves. Similarly, if the move occurred between Waves 7 and 8 and all waves are linked, then it would be more cost efficient to change the Wave 8 identifiers to match the corresponding fields in the early waves.

If the ultimate goal is to produce longitudinal units for analysis, the choic of options 1 and 2 as described are not satisfactory. This is because bot options obscure household, family, and program unit composition for some montr of the survey. Note that the current method of identifying households in ar month is the concatenation of SU-ID and address identifier (ADDID), the latte of which varies by month. This deficiency can be corrected in two ways. Or should use option 3, being careful to set the month-in-sample indicators (PP-MIS so that the individual is not over-represented in any particular month. Th second method is to retain monthly household identifiers in the linked fil which do not rely on the sample unit identifier (SU-ID). This can be handle by retaining the original value of SUSEQNUM from each wave and using that together with ADDID, to denote monthly household composition, being careful t match the corresponding waves and months. All other units such as families ar currently identified within household units, so that the preceding discussio applies if longitudinal units other than households are desired.

Note that, for any of the first three options, the user will need to combin information across two records for sample unit movers who moved within the wave in order to have one complete record for the individual. The questionnair data should be extracted entirely from the record of the person with the ne identifier. The control card data and weights should be extracted from th record with the old identifier before the move and the other record after th move.[8]

LINKING NONSUCCESSIVE WAVES

In order to merge topical information in one wave to core or topical informatic in another wave, it may be necessary to link nonsuccessive waves of SIPP. Le these waves be N and N+K. The process is the same as that described for linkir successive waves, i.e., person-level extracts are created, sorted, and the merged using SU-ID, PP-ENTRY, and PP-PNUM. However, the possible outcomes fo each individual surveyed are more complex then those described in the previou sections. The complexity arises because an individual can fall into multipl groupings over the period between Waves N and N+K. Hence, the user needs t construct independent controls for the merge by combining information from th User Note Series across the series of wave pairs N and N+1, N+1 and N+2, . . . N+K-1 and N+K.

To be more explicit, to determine the number of people whose identifier change between the two waves (N and N+K), examine the lists provided in the User Note distributed for all Waves N, N+1, . . . N+K. To determine all persons wh entered the sample after Wave N, examine the list of new sample members in eac successive wave starting with Wave N+1, and then delete from that list all suc persons listed as having left in subsequent waves through Wave N+K. To locat all persons who left the sample between Waves N and N+K, examine the list c persons who left between Waves N and N+1, and subtract from that all persor who reentered before Wave N+K and remained in the sample. The list of persor

---

[8]When the merge occurs within a wave, the user needs to be aware of the differ ent ways demographic and economic variables are handled. See chapter 4 for a explanation of that process.

who will not match between Waves N and N+K are the persons on these two lists, i.e., those who entered after Wave N and those who left before Wave N+1.

Aside from the complexity of the possible outcomes for individuals over time, one cautionary measure should be taken in the merge of topical information between waves (successive or not). It is possible that, due to a change in circumstances between waves, information collected in one wave may not be appropriate in another wave. For example, if the user decides to link child care expenses in Wave 5 of the 1984 Panel to persons in Wave 1, it would not be appropriate to merge positive child care expenses from Wave 5 to Wave 1 when the first child was born between the two waves. Hence, even when individual records correctly match between two waves, the user should do some exploratory analysis to locate observations for whom the topical information should not be used directly. Those cases should be treated in the same manner as individuals interviewed in Wave N but not in Wave N+K (or vice versa depending on the circumstances). That is, the topical information should either be flagged as not available, imputed, or the case should be deleted from the study with the appropriate weight adjustment to the other records.

REFERENCES

CITRO, C., D. HERNANDEZ, and R. A. HERRIOT. "Longitudinal Household Concepts in SIPP: Preliminary Results," Second Annual Research Conference Proceedings, March 23-26, 1986. Washington, D.C.: U.S. Bureau of the Census, 1986

KALTON, G. et al. "The Treatment of Person-Wave Nonresponse in Longitudinal Surveys," SIPP Working Paper Series No. 8704. Washington, D.C.: U.S. Bureau of the Census, 1987.

McMILLEN, D. B. and R. A. HERRIOT. "Toward a Longitudinal Definition of Households," Journal of Economic and Social Measurement (Formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4 (December 1985), 361.

CHAPTER 7 - ASSESSING RELIABILITY OF SIPP DATA

Data found in the Survey of Income and Program Participation (SIPP) publications or in user tabulations from the SIPP microdata are estimates based on the weighted counts from the survey. These estimates are subject to errors of two different kinds: sampling error, or errors due to the fact that the results from the SIPP sample may differ from those that might have been obtained if the entire population had been surveyed (i.e., if a census had been taken); and nonsampling error, or errors due to undercoverage and nonresponse, and errors made during data collection and processing.

Analysts should use standard errors and confidence intervals described in this chapter in deriving estimates from SIPP in order to account for the first type of error. The magnitude of the second type of error has not yet been fully evaluated, but several research papers describe the sources of nonsampling error in SIPP (Kalton, McMillen, and Kasprzyk, 1986; King, Petroni, and Singh, 1987).

SAMPLING ERROR

Measuring Sampling Variability: Standard Errors and Their Uses

Standard errors indicate the magnitude of the sampling variability. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors, for the most part, measure the variations that occurred by chance because a sample was surveyed instead of the entire population.

1. Constructing Confidence Intervals

The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same general conditions and using the same sample design, and if an estimate, its standard error, and confidence intervals were calculated from each sample, then:

1) Approximately 68 percent of the confidence intervals with a range of from one standard error below to one standard error above each estimate would include the average result of all possible samples.

2) Approximately 90 percent of the confidence intervals with a range from 1.6 standard errors below to 1.6 standard errors above the estimate would include the average result of all possible samples.

3) Approximately 95 percent of the confidence intervals with a range of two standard errors below to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derivable from all possible samples may or may not be contained in any particular computed interval. However, for a particular sample, one can say with a specified level of confidence that the confidence interval computed from that sample includes the value we are trying to estimate, i.e., the average estimate derivable from all possible samples.

2. Testing Hypotheses

Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses tested are (1) that the population parameters are identical versus (2) that they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the parameters are different when, in fact, they are identical.

To perform the most common test, let X_A and X_B be sample estimates of two parameters of interest. A subsequent section explains how to derive a standard error on the difference X_A-X_B. Let that standard error be S_{DIFF}. Compute the ratio $R=(X_A-X_B)/S_{DIFF}$. If this ratio is between -2 and +2, no conclusion about the parameters is justified at the 5-percent significance level. If, on the other hand, this ratio is less than -2 or greater than +2, the observed difference is significant at the 5-percent level.

In this event, it is a commonly accepted practice to say that the parameters are different. Of course, sometimes this conclusion will be wrong. When the parameters are, in fact, the same, there is a 5-percent chance of concluding that they are different.

Calculating Standard Errors for SIPP

There are two ways for users to compute a standard error for SIPP estimates. One method is to compute variances directly using pseudo-half-sample and pseudo-stratum codes contained on each household record.[1] The other method involves calculating generalized standard errors using simple charts and formulas found in published reports and the microdata documentation.

1. Computing Variances Directly

Pseudo-half-sample and pseudo-stratum codes (assigned in such a way as to avoid any disclosure risk) are included on the file to enable computation of variances by methods such as balanced repeated replications, Taylor Series linearization, and Jack-Knife repeated replications (Wolter, 1985; Kalton, 1977; Cochran, 1977). These methods may be used if the user cannot use the generalized standard errors, as in computing the variance of a correlation coefficient between, say, interest income and dividend income. A number of statistical software packages provide procedures for using half-sample codes (see Wolter, 1985).

Variances computed directly from SIPP public-use microdata will vary from SIPP variances estimated by the Census Bureau. These differences result from the use of artificial stratum codes on the public-use file, whereas the Census Bureau has access to the actual stratum identifiers. Actual stratum codes are withheld from the public-use microdata so as to avoid identifying geographic areas so small that they risk disclosure of confidential information.

---

[1]The relevant fields in the microdata files are H*-HSC and H*-STRAT.

Even though these are artificial stratum codes, the variance estimates are expected to be similar to those produced by the Bureau using the real stratum codes. Using the computation of direct variances method is involved, may be expensive, and, of course, is available only to users of SIPP microdata.

2. Generalized Standard Errors

To derive standard errors that are applicable to a wide variety of statistics and can be prepared at a moderate cost, a number of approximations are required. Most of the SIPP statistics have greater variance than those obtained through a simple random sample because clusters of living quarters are sampled for SIPP.

Two parameters, denoted "a" and "b," have been developed to calculate variances for each type of characteristic. These "a" and "b" parameters are used in estimating standard errors of survey estimates which are referred to as generalized standard errors. Parameters for Waves 1-9 of the 1984 Panel are given in figure 7.1. For some topical modules, special parameters may be issued as well. If issued, these parameters will be given in the technical documentation for the appropriate wave.

All statistics do not have the same variance behavior; "a" and "b" parameters were computed for groups of statistics with similar variance behavior. The parameters were computed directly for Waves 1 through 4 from SIPP third quarter 1983 data. These parameters were then adjusted to account for attrition and sample reduction.

The "a" and "b" parameters may be used to approximate the standard error for estimated numbers and percentages. Because the actual increase in variance was not identical for all statistics within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error, rather than the precise standard error for any specific statistic. That is why they are referred to as generalized standard errors.

Examples Using Generalized Standard Errors

The following sections explain the use of "a" and "b" parameters, found in figure 7.1, for computing a standard error and the corresponding confidence intervals. The values used are those from figure 7.1 which pertain to the early waves of the 1984 Panel.

1. Standard Error of Total

The formula for computing the standard error for a total is

$$s = \sqrt{ax^2 + bx} \qquad (1)$$

where "a" and "b" are the parameters associated with the estimate for the particular reference period and "x" is the weighted estimate.

Based on a tabulation from the SIPP survey data, you would find that there were 19,000,000 households with a monthly income of $3,000 and over during the fourth quarter of 1983. Suppose you want to develop a 95-percent confidence interval so you can assess how precise 19,000,000 is as an estimate.

Step 1:

Determine the appropriate "a" and "b" parameters by looking them up in figure 7.1. Since we are dealing with income data for all households, the "a" and "b" parameters are -.0000764 and 6766 (row 13, figure 7.1).

Step 2:

Enter these figures in the above formula.

$$s = \sqrt{ax^2 + bx}$$

$$= \sqrt{(-.0000764) \times (19{,}000{,}000)^2 + (6766 \times 19{,}000{,}000)}$$

$$= 317{,}763 \approx 318{,}000$$

where 19,000,000 is the estimate, and -.0000764 and 6766 are the "a" and "b" parameters. The resulting standard error (rounded to thousands) is 318,000.

Step 3:

To determine the 95-percent confidence interval of the estimate, multiply the standard error by 2, yielding 636,000. The lower bound of the confidence interval is then 19,000,000 minus 636,000 or roughly 18.4 million, and the upper bound is 19,000,000 plus 636,000 or roughly 19.6 million.

Thus we can conclude with 95-percent confidence that the average estimate derived from all possible samples lies within the interval of 18.4 million to 19.6 million.

2. Compensating for the Use of Partial Samples

The foregoing example assumes you are working with the full SIPP sample, as will normally be the case with SIPP reports and user tabulations. But if you are making a tabulation from SIPP microdata for a reference month for which you do not have data for all rotation groups, you must weight up the estimate by an appropriate factor to compensate for the smaller sample size; you must similarly adjust the estimates of the standard error.

FIGURE 7.1 - SIPP 1984 GENERALIZED VARIANCE PARAMETERS FOR ESTIMATED NUMBERS AND PERCENTAGES OF HOUSEHOLDS AND PERSONS

Waves 1-4 of the 1984 Panel

	Population Estimates[1]	Basic Parameters a	b
1.	16+ total persons: program participation and benefits, poverty (3)	-0.0000942	16059
2.	As above for 16+ total males	-0.0001984	16059
3.	As above for 16+ total females	-0.0001796	16059
4.	16+ total persons: income recipiency, labor force (4)	-0.0000321	5475
5.	As above for 16+ total males	-0.0000676	5475
6.	As above for 16+ total females	-0.0000612	5475
7.	0+ total persons:[2] all items (5)	-0.0000863	19911
8.	As above for total males	-0.0001786	19911
9.	As above for total females	-0.0001672	19911
10.	Black persons: poverty (1)	-0.0004930	13698
11.	As above for Black males	-0.0010522	13698
12.	As above for Black females	-0.0009274	13698
13.	Black persons: all other items (2)	-0.0002669	7366
14.	As above for Black males	-0.0005736	7366
15.	As above for Black females	-0.0004992	7366
16.	Total households: all items	-0.0000764	6766
17.	Black households: all items	-0.0004661	4675

FIGURE 7.1 - CONTINUED

Wave 5 of the 1984 Panel

	Population Estimates[1/]	Basic Parameters a	b
1.	16+ total persons: program participation and benefits, poverty (3)	-0.0001030	17539
2.	As above for 16+ total males	-0.0002167	17539
3.	As above for 16+ total females	-0.0001962	17539
4.	16+ total persons: income recipiency, labor force (4)	-0.0000351	5980
5.	As above for 16+ total males	-0.0000739	5980
6.	As above for 16+ total females	-0.0000669	5980
7.	0+ total persons:[2/] all items (5)	-0.0000943	21746
8.	As above for total males	-0.0001951	21746
9.	As above for total females	-0.0001827	21746
10.	Black Persons: poverty (1)	-0.0005422	14960
11.	As above for Black males	-0.0011652	14960
12.	As above for Black females	-0.0010140	14960
13.	Black persons: all other items (2)	-0.0002916	8045
14.	As above for Black males	-0.0006266	8045
15.	As above for Black females	-0.0005453	8045
16.	Total households: all items	-0.0000835	7390
17.	Black households: all items	-0.0005091	5106

FIGURE 7.1 - CONTINUED

Waves 6, 7, and 9 of the 1984 Panel

Population Estimates[1]	Basic Parameters a	b
1. 16+ total persons: program participation and benefits, poverty (3)	-0.0001144	20370
2. As above for 16+ total males	-0.0002404	20370
3. As above for 16+ total females	-0.0002182	20370
4. 16+ total persons: income recipiency, labor force (4)	0.0000390	6944
5. As above for 16+ total males	0.0000819	6944
6. As above for 16+ total females	0.0000744	6944
7. 0+ total persons:[2] all items (5)	-0.0001082	25255
8. As above for total males	-0.0002233	25255
9. As above for total females	-0.0002097	25255
10. Black persons: poverty (1)	-0.0006186	17372
11. As above for Black males	-0.0013259	17372
12. As above for Black females	-0.0011595	17372
13. Black persons: all other items (2)	-0.0003327	9343
14. As above for Black males	-0.0007131	9343
15. As above for Black females	-0.0006236	9343
16. Total households: all items	-0.0000993	8582
17. Black households: all items	-0.0006246	5929

FIGURE 7.1 - CONTINUED

Wave 8 of the 1984 Panel

Population Estimates[1]	Basic Parameters a	b
1. 16+ total persons: program participation and benefits, poverty (3)	-0.0001468	26141
2. As above for 16+ total males	-0.0003085	26141
3. As above for 16+ total females	-0.0002800	26141
4. 16+ total persons: income recipiency, labor force (4)	0.0000500	8912
5. As above for 16+ total males	0.0001051	8912
6. As above for 16+ total females	0.0000955	8912
7. 0+ total persons:[2] all items (5)	-0.0001389	32411
8. As above for total males	-0.0002866	32411
9. As above for total females	-0.0002691	32411
10. Black persons: poverty (1)	-0.0002583	22297
11. As above for Black males	-0.0017018	22297
12. As above for Black females	-0.0014883	22297
13. Black persons: all other items (2)	-0.0004270	11990
14. As above for Black males	-0.0009151	11990
15. As above for Black females	-0.0008003	11990
16. Total households: all items	-0.0001274	11013
17. Black households: all items	-0.0008016	7610

[1] Numbers in parentheses denote which parameter to use when cross-tabulating two characteristics. Use the characteristics with the minimum value.

[2] For example, use these parameters for retirement and pension tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

When you are working with fewer than all four rotation groups, the formula becomes

$$s = \sqrt{ax^2 + bx} \cdot \sqrt{f} \qquad (2)$$

where the first part of the expression is the same as before, and "f" is a factor compensating for sample size. In producing monthly estimates with one panel, the factor "f" is 4.0 when working with one rotation group, 2.0 with two rotation groups, and 1.33 with three rotation groups.[2] In other words, when the estimate is weighted up by a factor, the standard error must be multiplied by the square root of the same factor.

For example, the standard error in the above example was 318,000. If we were working with data for November 1983, using only the last two rotation groups in Wave 1 of the 1984 Panel, our initial estimate using the weights on the microdata file might have been 9,500,000. To compensate for the two missing rotation groups, we would apply the factor of 2.0, and thereby double our estimate to 19,000,000. The same factor would enter into the formula in equation (2) to give

$$s = 317{,}763 \times \sqrt{2.0} = 449{,}385 \simeq 449{,}000$$

as the standard error of an estimated 19,000,000 based on two rotation groups instead of four. The confidence interval is then determined in the same way, using this revised standard error.

Wave 1 represents a special case because there are 3 reference months at the start of the survey when the survey did not yet cover all four rotation groups. For example, in the 1984 Panel, only one rotation group has data for June 1983, two for July 1983, and three for August 1983. The first SIPP report (Current Population Reports, Series P-70, No. 1) included data for the third quarter 1983. For that period of partial coverage, a factor of 1.22, presented in chapter 5, is appropriate. If Wave 1 data alone had been used to estimate the fourth quarter example described above, the factor would be 1.85. Of course, Wave 2 supplies the missing rotation groups for that quarter so Wave 1 and Wave 2 files could be used together, thus creating the estimates from the full sample. In that case, no factor adjustment is needed.

The factors associated with the metropolitan area subsample (presented in chapter 5) are generally very close to 1.0, so they may be ignored in calculating variances for metropolitan summaries.

---

[2]Other factors associated with calendar quarter estimates were discussed in chapter 5.

3. Standard Error of a Percent

Computing the standard error and confidence interval for a percent follows a similar procedure. The formula for the generalized standard error of a percent is

$$s = \sqrt{\left(\frac{b}{y}\right) p(100-p)} \cdot \sqrt{f} \qquad (3)$$

where

y = the base of the percent (use weighted estimate); i.e., the size of the subclass of interest,

p = the percentage of persons, families, or households possessing the characteristic of interest,

b = the "b" parameter for the numerator of p, and

f = the factor to adjust for missing rotation groups if necessary.

Note that the "a" parameter is not used.

Suppose we find that of the households in Wave 1 who had a mean monthly income of $3,000 and over in the third quarter of 1983, 8,916,000 (8.6 percent) were Black. To construct a 95-percent confidence interval, follow the steps shown below.

Step 1:

Examine the "b" parameter in figure 7.1 for both total and Black households to determine the larger of the two. In this case, the "b" parameter for total households, 6766, is larger.

The "f" factor from chapter 5 (figure 5.9) that is applied to the base parameters to adjust for incomplete data is 1.22, applicable to third quarter data.

Step 2:

Entering the values into the formula in equation (3):

$$s = \sqrt{\left(\frac{6766}{8,916,000}\right) (8.6)(100-8.6)} \cdot \sqrt{1.22}$$

provides us with a standard error of 0.85 percentage points.

Step 3:

Multiplying the standard error by 2 and adding and subtracting this quantity from the estimate of 8.6 percent provides a 95-percent confidence interval of 6.9 to 10.3 percent.

4. Standard Error of a Difference

The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2 - 2rs_xs_y} \qquad (4)$$

where "s_x" and "s_y" are the standard errors of the estimates "x" and "y." The estimates can be numbers, percents, ratios, etc. The correlation between "x" and "y" is denoted by the correlation coefficient "r."[3] Figure 7.2 presents the correlation coefficients "r" for comparisons between months and between quarters. For other types of comparisons, assume "r" equals zero if it is believed that the value of one variable does not give a strong indication of the value of the other variable. If "r" is really positive, then this assumption will lead to overestimates of the true standard error. If "r" is negative, the result will be an underestimate of the actual standard error.

As an illustration, SIPP estimates show that the number of persons aged 35 to 44 years in nonfarm households with mean monthly household cash income over $4,000 during the third quarter of 1983 was 5,313,000; the number of those aged 25 to 34 years was 4,353,000, an estimated difference of 960,000. Using the Wave 1 parameters ("a"=-.0000321, "b"=5475, and "f"=1.22) in equation (2), the standard errors of the estimates for each age group are 185,422 and 168,324, respectively. It is reasonable to assume that these two estimates are not highly correlated. Therefore, the standard error of the estimated difference of 960,000 is

$$\sqrt{(185,426)^2 + (168,326)^2} = 250,432 \approx 250,000$$

[3]The correlation coefficient measures the extent to which the value of one variable gives an indication of the value of another variable. An example of a positive correlation is that between food stamp and AFDC recipiency. Food stamp and bond income recipiency are variables possessing a negative correlation. Another example of variables with positive correlation occurs when it is desired to measure the difference in a variable between two months or quarters.

5. Standard Error of a Mean

A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family, or household. For example, it could be the average monthly household income of females aged 25 to 34. The standard error of a mean can be approximated by the formula below. Because of the approximations used in developing the formula, an estimate of the standard error of the mean obtained from that formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean ($\bar{x}$) is

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2} \cdot \sqrt{f} \qquad (5)$$

where "y" is the size of the base, "s^2" is the estimated population variance of "x," "b" is the parameter associated with the particular type of item, and "f" is the adjustment factor.

The estimated population variance, "s^2," is given by formula (6):

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2 \qquad (6)$$

where there are n persons with the item of interest; "w_i" is the final weight for person i; and "x_i" is the value of the estimate for person i. In other words, this computation must be made from the individual microdata.

If the calculation of s^2 using formula (6) on the microdata is not practical and a frequency distribution is available as well as the mean, then formula (7) may be used instead:

$$s^2 = \sum_{i=1}^{c} p_i x_i^2 - \bar{x}^2 \qquad (7)$$

where each person (or other unit of analysis) is in one of c groups (e.g., income categories within an income distribution); the "p_i's" are the estimated proportions of responses within each group; the "x_i's" are the mid-

points of each group. If group c is open-ended. i.e., there is no obvious value for the midpoint "x_1," then an approximate value to use as the midpoint is

$$x_c = \left(\frac{3}{2}\right) Z_{c-1} \qquad (8)$$

where Z_{c-1} is the lower boundary of the group (e.g., $5,000 in the category $5,000 or more). If an open-ended group c does exist, the approximation could easily be bad. (See the section on top-coding in chapter 5.) To reduce this danger, use a large number of intervals so as to keep c and Z_{c-1} large.

6. Standard Error of a Mean Number of Persons with Characteristic Per Family or Household

Mean values for persons in families or households may be calculated as the ratio of two numbers. The denominator, y, represents a count of families or households of a certain class; the numerator, x, represents a count of persons with the characteristic under consideration who are members of these families or households. For example, the mean number of children per family with children is calculated as

$$\frac{x}{y} = \frac{\text{total number of children in families}}{\text{total number of families with children}}$$

For means of this kind, the standard error is approximated by the following formula:

$$s\left(\frac{x}{y}\right) = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left[\left(\frac{s_x}{x}\right)^2 - 2r\left(\frac{s_x}{x}\right)\left(\frac{s_y}{y}\right)\right]\right]} \qquad (9)$$

The standard error of the estimated number of families or households is "s_y"; the standard error of the estimated number of persons with the characteristic is "s_x." In the formula, "r" represents the correlation coefficient between the numerator and the denominator of the estimate. If at least one member of each family or household in the class possesses the characteristic of interest, then use 0.7 as an estimate of "r." If, on the other hand, it is possible that no member of a family or household has the characteristic, then use 0. In the example, you would use 0.7 for the average number of persons per family, but 0 for the average number of teenagers per family. Note that 0.7 is only a crude approximation; better estimates can be obtained using half-sample codes.

FIGURE 7.2 - CORRELATIONS FOR MONTHLY AND QUARTERLY COMPARISONS

Wave 1 Estimates	Total income, wage income and similar types of income	Program participation income, nonincome, labor force
Jun-Jul, Nov-Dec 1983	0.57	0.35
Jul-Aug, Oct-Nov 1983	0.65	0.41
Aug-Sep, Sep-Oct 1983	0.69	0.43
Jun-Aug, Oct-Dec 1983	0.43	0.26
Jul-Sep, Sep-Nov 1983	0.53	0.32
Aug-Oct 1983	0.50	0.30
Jun-Sep, Sep-Dec 1983	0.35	0.20
Jul-Oct, Aug-Nov 1983	0.29	0.16
Jun-Oct, Jul-Nov, Aug-Dec, Jun-Nov, Jul-Dec, Jun-Dec 1983	0.00	0.00
3rd Quarter-4th Quarter 1983	0.28	0.14
Wave 2 Estimates		
Oct-Nov 1983, Feb-Mar 1984	0.57	0.35
Nov-Dec 1983, Jan-Feb 1984	0.65	0.41
Dec 1983-Jan 1984	0.80	0.50
Oct-Dec 1983, Jan-Mar 1984	0.43	0.26
Nov 1983-Jan 1984, Dec 1983-Feb 1984	0.61	0.37
Oct 1983-Jan 1984, Dec 1983-Mar 1984	0.40	0.23
Nov 1983-Feb 1984	0.35	0.20
Oct 1983-Feb 1984, Nov 1983-Mar 1984 Oct 1983-Mar 1984	0.00	0.00
4th Quarter 1983-1st Quarter 1984	0.34	0.20

FIGURE 7.2 - CONTINUED

Wave 1 and 2 Combined Estimates	Total income, wage income and similar types of income	Program participation income, nonincome, labor force
Jun-Jul 1983, Feb-Mar 1984	0.57	0.35
Jul-Aug 1983, Jan-Feb 1984	0.65	0.41
Aug-Sep 1983, Dec 1983-Jan 1984	0.69	0.43
Sep-Oct, Oct-Nov,Nov-Dec 1983	0.80	0.50
Jun-Aug 1983, Jan-Mar 1984	0.43	0.26
Jul-Sep 1983, Dec 1983-Feb 1984	0.53	0.32
Aug-Oct 1983, Nov 1983-Jan 1984	0.65	0.39
Sep-Nov, Oct-Dec 1983	0.75	0.45
Jun-Sep 1983, Dec 1983-Mar 1984	0.35	0.20
Jul-Oct 1983, Nov 1983-Feb 1984	0.50	0.28
Aug-Nov 1983, Oct 1983-Jan 1984	0.61	0.35
Sep-Dec 1983	0.70	0.40
Jun-Oct 1983, Nov 1983-Mar 1984	0.33	0.18
Jul-Nov 1983, Oct 1983-Feb 1984	0.46	0.25
Aug-Dec 1983, Sep 1983-Jan 1984	0.56	0.30
Jun-Nov 1983, Oct 1983-Mar 1984	0.30	0.15
Jul-Dec 1983, Sep 1983-Feb 1984	0.42	0.21
Aug 1983-Jan 1984	0.60	0.30
Jun-Dec 1983, Sep 1983-Mar 1984	0.28	0.13
Jul 1983-Jan 1984, Aug 1983-Feb 1984	0.45	0.20
Jun 1983-Jan 1984, Aug 1983-Mar 1984	0.29	0.12
Jul 1983-Feb 1984	0.25	0.10
Jun 1983-Feb 1984, Jul 1983-Mar 1984	0.00	0.00
Jun 1983-Mar 1984		
3rd Quarter-4th Quarter 1983	0.63	0.36
4th Quarter 1983-1st Quarter 1984	0.51	0.29
3rd Quarter 1983-1st Quarter 1984	0.39	0.18

7. Standard Error of a Median

The median quantity of some item such as income for a given group of persons, families, or households is that quantity such that at least half the group has as much or more and at least half the group has as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. An approximate method for measuring the reliability of an estimated median is to determine a confidence interval (ranges that would include the average result of all possible samples with a known probability) about it. The following procedure may be used to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

Step 1:

Determine, using the formula for the standard error of an estimated percentage, the standard error of an estimate of 50 percent of the group. Example: The median monthly income for persons age 25 to 34 is $2,122. The size of the group is 39,533,000. Using formula (10), the standard error of 50 percent on a base of 39,533,000 is about 1.1 percentage points.

$$s_{(x,p)} = \sqrt{\frac{b}{x}\ p^{(100-p)}} \qquad (10)$$

Step 2:

Add to and subtract from 50 percent the standard error determined in step (1). Following step (2), the two percentages of interest are 48.9 and 51.1.

Step 3:

Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group owning more is equal to the smaller percentage found in step (2). This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group owning more is equal to the larger percentage found in step (2). This quantity will be the lower limit for the 68-percent confidence interval. By examining figure 7.3, we see that the percentage 48.9 falls in the income interval from $2,000 to $2,499. Thus:

A_1 = $2,000
A_2 = $2,500
N_1 = 21,595,000
N_2 = 15,466,000

Since the median is greater than $200, Pareto interpolation is indicated. So the upper bound of a 68-percent confidence interval for the median is

$$(\$2{,}000)\exp\left[\frac{\ln\left(\frac{(.489)(39{,}533{,}000)}{21{,}595{,}000}\right)}{\ln\left(\frac{\$2{,}500}{\$2{,}000}\right)}\right] = \$2{,}154$$

Also by examining figure 7.3, we see that the percentage of 51.1 falls in the same income interval. Thus, A_1, A_2, N_1, and N_2 are the same. So the lower bound of a 68-percent confidence interval for the median is

$$(\$2{,}000)\exp\left[\frac{\ln\left(\frac{(.511)(39{,}533{,}000)}{21{,}595{,}000}\right)}{\ln\left(\frac{\$2{,}500}{\$2{,}000}\right)}\right] = \$2{,}091$$

Step 4:

Divide the difference between the two quantities determined in step (3) by two to obtain the standard error of the median. Thus, the 68-percent confidence interval on the estimated median is from $2,091 to $2,154. An approximate standard error is

$$\frac{\$2{,}154 - \$2{,}091}{2} = \$32$$

To perform step (3), it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. For this report, we recommend Pareto interpolation for any point in a monthly income interval greater than $200, and linear interpolation otherwise.

Interpolation is used as follows. The quantity of the item such that "p" percent own more is

$$X_{pN} = A_1 \exp \left[\overline{\text{Ln}\left(\frac{pN}{N_1}\right)} \overline{\text{Ln}\left(\frac{A_2}{A_1}\right)} \Big/ \overline{\text{Ln}\left(\frac{N_2}{N_1}\right)} \right] \qquad (11)$$

if Pareto interpolation is indicated and

$$X_{pN} = \frac{N_1 - pN}{N_1 - N_2} (A_2 - A_1) + A_1 \qquad (12)$$

if linear interpolation is indicated,

where

- N = size of the group,
- A_1 and A_2 = the quantities of the item which can be easily seen to be the lower and upper bounds, respectively, of the interval in which X_{pN} falls,
- N_1 and N_2 = the estimated number of group members owning more of the item than A_1 and A_2, respectively,
- exp = refers to the exponential function, and
- Ln = refers to the natural logarithm function.

It should be noted that a mathematically equivalent result is obtained by using common logarithms (base 10) and antilogarithms.

8. Standard Errors of Ratios of Means or Ratio of Medians

In this section, the correlation between the numerator and denominator, "r," is assumed to be zero, as is true for the ratio of means for disjoint demographic groups. So, the standard error for a ratio of means or medians is approximated by this formula:

$$s_{\left(\frac{x}{y}\right)} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]} \qquad (13)$$

Where x and y are the means or median and the standard errors of the two means or medians are s_x and s_y. If "r" is positive, as is true when comparing means

for the same demographic group, then this procedure will provide an overestimate of the standard error for the ratio of means and medians. If "r" is negative, as is true for mean earnings and welfare, then this procedure will provide an underestimate.

NONSAMPLING ERROR

Nonsampling errors in panel surveys can arise from many sources. A review of the sources of nonsampling error in SIPP is found in Kalton, McMillen, and Kasprzyk (1986) and King, Petroni, and Singh (1987). Several sources of nonsampling error are discussed briefly below. In Kalton, McMillen, and Kasprzyk (1986), nonsampling errors which arise from the major SIPP decision to employ a panel design, namely panel nonresponse and panel conditioning, are discussed, as well as other design decisions, such as the length of the reference period, respondent rules, following rules, and data collection mode. These design decisions are reviewed in light of their possible effects on nonsampling errors. The latter source (King, Petroni, and Singh, 1987) discusses a broader array of topics related to the sources of nonsampling error in the SIPP, including sampling frame construction, implementation of the sampling scheme, undercoverage, data collection procedures, measurement error, and nonresponse error. If interested in any of these topics, the reader should go to these documents.

The remainder of this chapter discusses nonresponse, both unit and item, and the use of imputation indicators on the public-use file.

Nonresponse and Imputation

As is true for all surveys, some information is not collected which in theory should have been. Addresses in the original sample could be missed or all residents of the sample could refuse to participate. Individuals in otherwise successfully interviewed households could refuse to participate (these are called Type Z nonrespondents). Furthermore, individuals could participate in the survey by responding to most but not all questions; some questions are left unanswered either because the answer is unknown or the person simply refused. This situation is referred to as item nonresponse.

Users of survey data have three options to compensate for nonresponse. First, responding households or individuals could be reweighted. Second, missing responses could be imputed based on information obtained from other respondents. Third, the incomplete cases can be analyzed directly using appropriate analytic tools as described in Little and Rubin (1987). In the SIPP microdata files noninterviews are dealt with by weighting up interviewed cases, and individual nonrespondents and item nonresponse are dealt with through imputation.[4]

The procedures employed to adjust the weights of respondent households is described in chapter 5. The procedures used to impute data for nonresponding individuals and item nonresponse will be described in subsequent editions of the Users' Guide.

---

[4]There was an exception for Wave 1 of the 1984 Panel. Interviewed households containing a Type Z nonrespondent were assigned zero weights in that wave, and other similar interviewed households had their weights adjusted accordingly.

In general, an imputation procedure enhances the usefulness of the data. It simplifies processing for the microdata user by eliminating "not reported" categories. By imputing a missing characteristic with that of someone similar in other key aspects, the user can work with a more complete data set. Statistically there are advantages and disadvantages to the use of imputed data. Users wishing to know more about this issue are referred to Madow, Olkin, and Rubin (1983); Kalton (1983); Kalton and Kasprzyk (1986); and Little (1986).

Noninterviewed and Ineligible Addresses

SIPP is a longitudinal survey of adults in an initial sample of interviewed addresses. Addresses to be included in the initial interviewing process for each panel are determined through sampling procedures similar to the CPS. Addresses to be included in subsequent rounds of interviews for each panel are determined by the original sample adults. (This aspect of the survey design is described more fully in chapter 2.)

A certain number of addresses selected into the SIPP sample fail to generate interviews while others are ruled ineligible for inclusion in the survey. In Wave 1 each address not successfully interviewed is classified by type of noninterview as follows:

Type A: A unit occupied by persons eligible for an interview but for whom no questionnaire is completed. The reasons for Type A noninterview include no one at home, occupants temporarily absent, occupants refused to give information, unable to locate, and other.

Type B: A unit not eligible at the time of interview for inclusion in the survey; these noninterviews include vacant units, units temporarily converted to a business, slated to be demolished or under construction, and units occupied solely by persons whose usual residence is elsewhere.

Type C: A unit not eligible at the time of interview for inclusion in the survey; these noninterviews include units that no longer exist because they were demolished, converted permanently to a business, or merged with another unit.

Addresses classified as Type B or C noninterview in the initial Wave 1 sample are ineligible for the survey. Type A addresses are eligible for the survey but due to the inability to obtain interviews, they are dropped. Adults in interviewed households in Wave 1 of each panel define the SIPP longitudinal sample.[5]

[5]This statement is true for the 1984, 1986, and subsequent panels. In the 1985 Panel, Type A noninterviews in Wave 1 could be converted and included in the sample if they were converted by Wave 2. Otherwise, they were dropped.

In subsequent waves of each panel, each noninterviewed address is classified as follows:[6]

Type A: A unit occupied by persons eligible for an interview but for whom no questionnaire is completed. The reasons for Type A noninterview include no one at home, occupants temporarily absent, occupants refused to give information, unable to locate, and other.

Type C: All original sample adults in the unit were out of scope, either through death or moving abroad; therefore, they were not followed.

Type D: Person or group of persons to be followed moved outside the interviewing radius or their address was unknown.

Figure 7.4 shows the number of addresses for which an interview was attempted in each Wave of the 1984 Panel by whether they were eligible for interview, whether an interview was conducted, and, if no interview was conducted, the type of noninterview. Of note is the Type A noninterview rate, as this rate is the most susceptible to reduction. Except for Type A noninterviews in Wave 1 which are excluded from the sample,[7] interviewers are encouraged to make repeated efforts to convert refusals or locate respondents. "New" Type A noninterviews represent the first time a Type A noninterview occurred at a particular address. "Old" Type A noninterviews represent unsuccessful attempts to convert a Type A noninterview from the previous wave. In general, two consecutive Type A noninterviews at a particular address will result in the unit being eliminated from the cross-sectional public-use files after the second noninterview. There were some exceptions to this in the 1984 and 1985 Panels.

Figure 7.4 also lists Type D noninterviews. These occur when one or more individuals move to a new address which cannot be located, or which is beyond 100 miles from a SIPP Primary Sampling Unit (PSU) and cannot be contacted by telephone. Note that this type of noninterview can apply to all residents of an interview address in the previous wave or to a subset of those residents.

A discussion of the SIPP sample loss and efforts to reduce it is found in Nelson, Bowie, and Walker (1987). One conclusion the authors reach is that the overall cumulative SIPP sample loss appears to be comparable to that which is observed by other longitudinal surveys. A discussion of nonresponse adjustment procedures used to compensate for survey nonresponse can be found in Chapman, Bailey, and Kasprzyk (1986).

---

[6]Aside from the noninterview codes noted, Type B noninterview codes were introduced in Wave 6 (of the 1984 Panel) and Wave 2 (of the 1985 Panel) when the Census Bureau began tracking original sample members who entered institutions. Type B is assigned when all members of an address entered an institution.

[7]There was an exception for Wave 1 of the 1984 Panel. Interviewed households containing a Type Z nonrespondent were assigned zero weights in that wave, and other similar interviewed households had their weights adjusted accordingly.

FIGURE 7.4 - SAMPLE SIZES IN ADDRESSES BY WAVE AND INTERVIEW STATUS, 1984 PANEL

Wave	Total	Ineligible			Eligible								
		Total	Out of Scope	No Sample Persons Left	Total	Interview	Noninterview						
							Total	Old A's	New A's	Old D's	New D's	Offshoots of Old A's and D's	Rate
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	26024	5127	5127	0	20897	19878	1019	0	1019	0	0	0	4.9
2	16876	833	97	736	16043	14532	1511	762	570	0	159	20	9.4
3	23282	1407	130	1277	21875	19191	2684	1498	721	159	244	62	12.3
4	23911	1542	121	1421	22369	18932	3437	2093	599	403	219	123	15.4
5*	N/A	N/A	N/A	N/A	20544	16978	3566	2344	397	561	128	136	17.4
6*	N/A	N/A	N/A	N/A	18683	15061	3622	2419	335	628	90	150	19.4
7	N/A	N/A	N/A	N/A	18863	14902	3961	2694	214	718	151	184	21.0
8	N/A	N/A	N/A	N/A	14225	11096	3129	2104	108	640	175	152	22.0

Notes by Column:

(1) Wave 2 and Wave 8 were only 3 months long. Thus the small sample sizes. Also, there was a sample cut in Wave 5.

(2), (3), (4), (5) Available only for the first four waves. (4) Out-of-scope addresses are those that either do not exist, are not fit for habitation, or are vacant. (5) Addresses not occupied by original sample persons.

(9), (10) Type A noninterviews are addresses for which no interview was obtained due to refusal or inability to find someone at home. Old A's are Type A noninterviews that carry over from the previous wave.

(11), (12) Type D noninterviews result when a mover cannot be followed. Old D's are Type D noninterviews that carry over from the previous wave.

(13) People move out of old A's and D's and set up new households. Each of these new households is eligible for interview. However, it is generally impossible to find them. This column is projected from growth rates among interviewed households.

(14) The percent of eligible households that were not interviewed.

*Beginning in Wave 5, rotation 2, sample was cut by 17.8 percent. This continued through rotation 1 of Wave 6.

Inclusion of Imputation Flags

The initial evaluation of the quality of the data from SIPP shows improvements in the accuracy and completeness of the data on income and program participation over that obtained from the March CPS. For example, for the third quarter of 1983, SIPP item nonresponse rates ranged from a low of about 3 percent for questions about Aid to Families with Dependent Children and food stamp allotments, to about 13 percent for those concerning self-employment income. These rates contrast sharply with the higher nonresponse rates from the March CPS. The rates for CPS range from a low of 9 percent for food stamp allotments to 24 percent for self-employment income. The nonresponse rate for monthly wage and salary income overall averaged about 6.2 percent for the initial SIPP interviews, compared to 18.9 for the March 1985 CPS. A more complete summary of the item nonresponse rates in SIPP and the improvement in response rates over the March CPS can be found in the Current Population Reports, Series P-70, Nos. 1-6.

If the characteristics of nonrespondents are systematically different from the characteristics of respondents, as may well be the case for income variables, then it is possible that the imputation system masks certain biases due to nonresponse. For this reason, the SIPP microdata files include flags for many data items which allow the user to discriminate between those responses which were actually reported and those entries which were supplied through imputation. The flags denoting imputation for item nonresponse appear at the end of the household, person, and income records in the SIPP complex microdata file, and at the end of appropriate sections within the records of the rectangular file, generally corresponding one-for-one with specific data items. Individuals for whom the entire record was imputed are Type Z nonrespondents (PP-INTWV = 3 or 4) and those persons flagged as "not-in-sample" in the interview month and flagged as "in-sample" in at least one of the 4 reference months.[8] (The variable PP-MIS denotes "in-sample."[9])

In the example in figure 7.5, the data item for earned income received from a particular job in a particular month is shown on the top half (WS1-AMT4 is the value after imputation, WS1-2032 is the field in which reported value should occur). A sample value of 2000 is illustrated for the final amount, WS1-AMT4, i.e., $2000 of income last month.

---

[8]Type A or D individuals can be considered "not in sample" in the interview month and "in sample" in at least 1 of the 4 reference months. Type D individuals could have moved out-of-scope after the middle of the first reference month. In this case, they were part of the sample and had missing data during the early part of the wave. Also, individuals in Type A addresses could have split off from other groups which were successfully interviewed in the wave of the split. If the move occurred after the middle of the first reference period, they are considered to have missing data in the months before the move.

[9]Note that these individuals have the item imputation flags from the donor records assigned along with all of the other person data. These imputation flags should be ignored.

FIGURE 7.5 - ILLUSTRATION OF AN IMPUTATION FLAG ON WAVE 3, 1984 PANEL

Data Dictionary				Sample Values
		(Wage and Salary Record)		
Sample Data Item				
D WS1-AMT4	5	3274		2000
What is the dollar amount of the earnings from the job for month 4. Range = 0,33332.				
U Persons 15 years old and older				
V -9. Not in universe				
D WS1-2032	5	3298		2000
What was the total amount of pay that ... received before deductions on this job last month (month 4). Range = -9,33332.				
U Persons 15 years old and older				
V -9. Not in universe				
0. None				
Corresponding Imputation Flag				
D WS1CAL01	1	3326		1
Field 'WS1-2032' was calculated				
V 0. Not calculated				
1. Imputed input				
2. No imputed input				

Its corresponding imputation flag is shown on the bottom half. Note that the description of the imputation flag cites the field name for the originally reported item, WS1-2032. The sample value of "1" in the imputation flag indicates that the original respondent failed to answer the corresponding question, or the entry supplied was unusable for some reason; therefore, the information in the data item above was imputed from that of another person.

In examining only the final income amounts, one would not know that the $2,000 was imputed rather than actually reported by the individual. Only by cross-tabulating income by imputation status can one recognize an imputed income. Note that, except for wages and salaries and self-employment, imputation flags are either "0" (not imputed) or "1" (imputed) and appear in reverse order from the recoded amount fields.[10] Hence, imputed amounts can be easily determined by multiplying the amount by the corresponding imputation flag.

Unedited Values

There are a number of demographic characteristics from the control card which should not require imputation, but may need to be edited for consistency with other information from the household. In these cases there are no imputation flags, but the file includes both the edited value and the value prior to computer editing, referred to as preedited or unedited. These items are identified by a "U" at the start of the 8-character mnemonic identifying variables in the data dictionary. To detect whether a particular edit had any impact on the data, compare a given data item with its preedited or unedited counterpart.

Uses of Imputation Flags

An analyst can use imputation flags or unedited items in several different ways.

- By computing the rate of imputation one can evaluate the quality of certain data items. For instance, one could find out whether persons receiving aid from the government are less likely to report their other sources of income than persons not participating in such programs.

- Imputation flags allow characteristics of nonrespondents to be studied. Do nonrespondents tend to be younger or older, for example, than the rest of the population?

- One can exclude imputed data from cross-tabulations that might be sensitive to the imputation procedures. For instance, in comparing the earnings of doctors and dentists, missing income on a doctor's or dentist's record would be imputed from a pool of possible donors which would include a much broader range of professional occupations. Thus, to make sure

[10] The flags appear in the order in which the monthly income amounts were collected (1 month before interview, 2 months before interview, etc.). The recoded amounts, on the other hand, appear in chronological sequence. Note the technical documentation for the rectangular core files for the 1984 Panel erroneously states that the imputation flags appear in the reverse order, i.e., 4 months before interview, 3 months before interview, etc.

you are strictly comparing doctors' incomes with dentists' incomes, it would be appropriate to exclude all cases with either occupation or income imputed.

- Finally, one could compare the results with and without imputed data, to assess the effects of imputation on the results of the study.

REFERENCES

CHAPMAN, D., L. BAILEY, and D. KASPRZYK. "Nonresponse Adjustment Procedures at the U.S. Bureau of the Census," Survey Methodology, Vol. 12, No. 2 (1986), pp. 161-180.

COCHRAN, W. G. Sampling Techniques, 3rd Ed. New York: John Wiley and Sons, 1977.

KALTON, G. Compensating for Missing Survey Data. Ann Arbor: Survey Research Center, University of Michigan, 1983.

KALTON, G. "Practical Methods for Estimating Survey Sampling Errors," Bulletin of the International Statistical Institute, 4713 (1977), pp. 495-514.

KALTON, G. and D. KASPRZYK. "The Treatment of Missing Survey Data," Survey Methodology, Vol. 12, No. 1 (1986), pp. 1-16.

KALTON, G., D. B. McMILLEN, and D. KASPRZYK. "Nonsampling Error Issues in the Survey of Income and Program Participation," Second Annual Census Bureau Research Conference Proceedings, March 23-26, 1986, pp. 147-164. Washington, D.C.: U.S. Bureau of the Census, 1986.

KING, K., R. PETRONI, and R. SINGH. "Quality Profile for the Survey of Income and Program Participation," SIPP Working Paper Series No. 8708. Washington, D.C.: U.S. Bureau of the Census, 1987.

LITTLE, R. J. A. and D. B. RUBIN. Statistical Analysis with Missing Data. Somerset, N.J.: John Wiley and Sons, Inc., 1987.

LITTLE, R. J. A. "Missing Data in Census Bureau Surveys," Second Annual Census Bureau Research Conference Proceedings, March 23-26, 1986. Washington, D.C.: U.S. Bureau of the Census, 1986.

MADOW, W. G., I. OLKIN, and D. B. RUBIN (editors). Incomplete Data in Sample Surveys, Vols. 1, 2, and 3. New York: Academic Press, 1983.

NELSON, D., C. BOWIE, and A. WALKER. "Survey of Income and Program Participation (SIPP) Sample Loss and the Efforts to Reduce It," Third Annual Census Bureau Research Conference Proceedings. Washington, D.C.: U.S. Bureau of Census, 1987.

WOLTER, K. Introduction to Variance Estimation. New York: Springer-Verlag, 1985.

APPENDIX A

ADDITIONAL REFERENCES

ADDITIONAL REFERENCES

BEEBOUT, H. "Estimated Work-Related Expenses for Food Stamp Households," Memorandum to U.S. Department of Agriculture, Food and Nutrition Service. Washington, D.C.: Mathematica Policy Research, Inc., March 1984.

BEEBOUT, H. "Reported Child Care Costs for Food Stamp Households on the ISDP," Memorandum to U.S. Department of Agriculture, Food and Nutrition Service. Washington, D.C.: Mathematica Policy Research, Inc., March 1984.

BICKEL, G. and M. MacDONALD. Assets of Low Income Households: New Findings on Food Stamp Participants and Nonparticipants. Washington, D.C.: Mathematica Policy Research, Inc., January 1981.

CARR, T., P. DOYLE, and I. LUBITZ. Turnover in Food Stamp Participation: A Preliminary Analysis. Washington, D.C.: Mathematica Policy Research, Inc., July 1984.

CITRO, C. and M. MANSER. "Examination of Alternative Income Distribution Measures Using the SIPP Panel Data," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

CODER, J. "Some Results From the 1979 ISDP Research Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

CZAJKA, J. Determinates of Participation in the Food Stamp Program: Spring 1979. Washington, D.C.: Mathematica Policy Research, Inc., November 1981.

CZAJKA, J. and C. CITRO. "Household Income and Poverty Statistics Under Alternative Measures of Household and Family Composition," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

DAVID, M. "Discussion of Papers: Czajka and Citro, Lane and Lininger," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

DAVID, M. "Discussion of Papers Presented at the 1980 ASA Meeting," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

DOYLE, P. (U.S. Department of Agriculture). Further Analysis of Serial Multiple Benefit Receipt. Washington, D.C.: Mathematica Policy Research, Inc., March 1985.

DOYLE, P. An Analysis of Intra-Year Income Receipt. Washington, D.C.: Mathematica Policy Research, Inc., July 1984.

DOYLE, P. and C. CITRO. The ISDP/RAMIS II System and Development. Washington, D.C.: Mathematica Policy Research, Inc., May 1984.

DUBNOFF, S., D. VAUGHAN, and C. LANCASTER. "Income Satisfaction Measures in Equivalence Scale Applications," Proceedings of the Section on Business and Economic Statistics, American Statistical Association, 1981.

FELDMAN, A., C. NELSON, and J. CODER. "Evaluation of Wage and Salary Income Reporting on the 1978 ISDP Test Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

FRAKER, T. and R. MOFFITT. The Effects of Food Stamp Participation on the Market Labor of Female Heads of Household. Washington, D.C.: Mathematica Policy Research, Inc., March 1985.

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 Economics Statistics, Vol. 2, No. 2, April 1984.

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," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

HEERINGA, S. "Small Area Estimation for SIPP," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

KALTON, G. "Handling Wave Nonresponse in Longitudinal Surveys," Proceedings of the First Annual Research Conference. Washington, D.C.: U.S. Bureau of the Census, 1985.

KALTON, G. and D. KASPRZYK. "Imputing for Missing Survey Responses," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

KALTON, G., D. KASPRZYK, and R. SANTOS. "Some Problems of Nonresponse and Nonresponse Adjustment in SIPP," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

KALTON, G. and L. KISH. "Some Efficient Random Imputation Methods," Communication in Statistics Theory and Methods, No. 13(16), 1984, pp. 1919-1939.

KALTON, G. and L. KISH. "Two Efficient Random Imputation Procedures," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

KASPRZYK, D. "Some Research Issues for the Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1983.

KASPRZYK, D. and C. LININGER. "The 1982 Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

KLEIN, B. "How the SIPP Can Respond to the Recommendations of the Levitan Commission," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

LANE, J. and J. OLSON. "Changes in Income and Program Participation During Six Months in a Low Income Sample," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

LANE, J. and C. LININGER. "Monthly Variation in Family Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

LININGER, C. "The Goals and Objectives of the Survey of Income and Program Participation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

LITTLE, R. "Discussion of Papers Presented at 1981 ASA," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

LONG, S. Turnover in the Food Stamp Program: A Methodological Review. Washington, D.C.: Mathematica Policy Research, Inc., May 1985.

LUBITZ, J. and T. CARR. Turnover in the Food Stamp Program in 1979: The Role of Trigger Events. Washington, D.C.: Mathematica Policy Research, Inc., February 1985.

MacDONALD, M. Serial Multiple Benefits and Monthly Income Adequacy. Washington, D.C.: Mathematica Policy Research, Inc., March 1985.

MacDONALD, M. Multiple Benefits and Income Adequacy: Impacts on Nutrition and Housing Benefits. Madison, Wisc.: Institute for Research on Poverty, August 1984.

MacDONALD, M. Multiple Benefits and Income Adequacy for Food Stamp Participant and Nonparticipant Households. Washington, D.C.: Mathematica Policy Research, Inc., February 1983.

McDONALD, M. "Multiple Benefits: Fitting the Pieces Together," Focus, Vol. 6, No. 3, Summer-Fall 1983 (Institute for Research on Poverty, University of Wisconsin-Madison).

MANSER, M. "Analyzing In-Kind Benefits Using the 1979 ISDP Panel Data," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

OLSON, J. and R. KLEIN. "Interviewers Perceptions of Reasons for Participation Refusal in a National Longitudinal Survey 1979-80," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

PEARL, R. "Discussion of Papers Presented at 1981 ASA," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

PEARL, R. and M. FRANKEL. "Composition of Personal Wealth of American Households," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

RODGERS, W. and E. DeVOL. "An Evaluation of Statistical Matching," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

SANTOS, R. "Effects of Imputation on Regression Coefficients," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

SLATER, C. "Discussion of Papers: Manser; Pearl and Frankel; and Vaughan, Lininger, and Whiteman," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

"SPECIAL ISSUE: The Survey of Income and Program Participation: 1984. (Papers presented at a conference sponsored by the Social Science Research Council) "Journal of Economic and Social Measurement (Formerly Review of Public Data Use), Vol. 13, Nos. 3 and 4, December 1985. (A summary of a follow-up seminar "Improving the Design of SIPP," in June 1985 is also included.)

SUTTON, G. "Measuring Welfare Eligibility Using New Data Sources," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

TEITEL, R. "Statistical Databases and Database Systems," Proceedings of the First Annual Research Conference. Washington, D.C.: U.S. Bureau of the Census, 1985.

U.S. DEPARTMENT OF COMMMERCE, BUREAU OF THE CENSUS. "Lifetime Work Experience and Its Effects on Earnings: Retrospective Data From the 1979 Income Survey Development Program," Current Population Reports, Special Studies, Series P-23, No. 136. Washington, D.C.: U.S. Government Printing Office, 1984.

U.S. DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS. "Wage and Salary Data From the Income Survey Development Program: 1979," Current Population Reports, Special Studies, Series P-23, No. 118. Washington, D.C.: U.S. Government Printing Office, 1982.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, SOCIAL SECURITY ADMINISTRATION. Income Survey Development Program, 1979 Research Panel Documentation. Washington, D.C.: National Technical Information Service, 1982. (This technical data file documentation contains an extensive bibliography on the ISDP.)

VAUGHAN, D. and C. LANCASTER. "Applying a Cardinal Measurement Model to Normative Assessments of Income: Synopsis of a Preliminary Look," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

VAUGHAN, D., C. LININGER, and R. KLEIN. "Differentiating Veterans' Pension and Compensation Income in the 1979 ISDP Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1983.

VAUGHAN, D., T. WHITEMAN, and C. LININGER. "The Quality of Income and Program Data in the 1979 ISDP Research Panel: Some Preliminary Findings," Review of Public Data Use, Vol. 12, 1984, pp. 107-131.

WERTHEIMER, R. "The Role of SIPP in Microsimulation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

WHITE, G. and H. HUANG. "Mover Follow-up Costs for the Income Survey Development Program," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1982.

WHITEMAN, T. "The Measurement of Nonfarm Self-Employment Income: The 1979 ISDP Research Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981.

YCAS, M. and D. KASPRZYK. "Operationalizing Alternative Disability Definitions in a Household Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

YCAS, M. and C. LININGER. "ISDP: A Review," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

APPENDIX B

WHAT'S AVAILABLE

WHAT'S AVAILABLE

July 1987

What's available from the Survey of Income and Program Participation?

A number of SIPP Working Papers and reports are now available at this time:

- Working Papers No. 8401, Nos. 8501-8507, Nos. 8601-8614, and Nos. 8701-8714.
- Compilation of papers presented in the SIPP sessions at the American Statistical Association meetings in August 1984, 1985, and 1986.
- "Economic Characteristics of Households in the United States" & "Household Economic Studies" (Current Population Reports Series P-70, Nos. 1-9).
- A special issue of the Journal of Economic and Social Measurement containing the proceedings of a conference sponsored by the National Science Foundation, the Social Science Research Council and the Census Bureau. The table of contents of this issue is attached for your information.
- "Food Stamp Research: Results from the Income Survey Development Program and the Promise of the Survey of Income and Program Participation," a report prepared by Mathematica Policy Research under a contract with the Food and Nutrition Service, U.S. Department of Agriculture.

SIPP public-use microdata files can be purchased through the Census Bureau's Data User Services Division (301-763-4100). (See Attachment 9 for a list of files available at this time.) Another way of accessing SIPP data is through the University of Wisconsin's Research Network and Data Center for SIPP. Call Alice Robbin at (608) 262-4574 or Martin David at (608) 262-3281 for more information.

For more current information on--

REPORTS: Contact SIPP Staff on 301-763-5784

TAPES: Contact Customer Services on 301-763-4100 (When ordering "complex" files, use the term "Relational.")

See the attached listing for a description of the available products from the Survey of Income and Program Participation (SIPP). Please note that Working Paper 8401 was revised in December 1985 to account for program additions and changes which took place in 1985. If you want to receive any of these products, put a check by the appropriate number and mail to:

Daniel Kasprzyk, Special Assistant
Population Division, Room 2025-3
Bureau of the Census
Washington, D.C. 20233

SIPP WORKING PAPERS

1984/1985 (See Att. 1)

___ 8401
___ 8501
___ 8502
___ 8503
___ 8504
___ 8505
___ 8506
___ 8507

1986 (Att. 2)

___ 8601
___ 8602
___ 8603
___ 8604
___ 8605
___ 8606
___ 8607
___ 8608
___ 8609
___ 8610
___ 8611
___ 8612
___ 8613
___ 8614

1987 (Att. 3)

___ 8701
___ 8702
___ 8703
___ 8704
___ 8705
___ 8706
___ 8707
___ 8708
___ 8709
___ 8710
___ 8711
___ 8712
___ 8713
___ 8714

COMPILATIONS OF ASA PAPERS

___ 1984 (Att. 4) ___ 1985 (Att. 5) ___ 1986 (Att. 6)

___ Journal of Economic & Social Measurement (Att. 7)

___ Food Stamp Research: Results of ISDP and the Promise of SIPP (Att. 8)

CURRENT POPULATION REPORTS, SERIES P-70

"Household Economic Studies," "Economic Characteristics of Households in the United States"

___ P-70, No. 1 (Third Quarter 1983)
___ P-70-83-4 (Fourth Quarter 1983)
___ P-70, No. 3 (First Quarter 1984)
___ P-70, No. 4 (Second Quarter 1984)
___ P-70, No. 5 (Third Quarter 1984)
___ P-70, No. 6 (Fourth Quarter 1984)

*"Household Economic Studies"

*With the release of the P-70, No. 7, report, the "Household Economic Studies" reports are no longer quarterly; they will be released intermittently throughout the year under different subject-matter titles.

** Available through GPO

**P-70, No. 7 (Household Wealth and Asset Ownership, 1984)
**P-70, No. 8 (Disability, Functional Limitation, and Health Insurance Coverage: 1984/85)
**P-70, No. 9 ("Who's Minding the Kids?" Child Care Arrangements: Winter 1984-85)

[ ] Please put me on the mailing list
[ ] I would like to be taken off the mailing list.

NAME: _____
ADDRESS: _____

Attachment 1

SIPP WORKING PAPERS - 1984/1985

8401 - (Update No. 1, Revised 12/85) "An Overview of the Survey of Income and Program Participation," by D. NELSON, D.B. MCMILLEN, and D. KASPRZYK, Census Bureau

---

8501 - "The Survey of Income and Program Participation: Uses and Applications," by K.S. SHORT, Census Bureau.

8502 - "Applications of a Matched File Linking the Bureau of the Census Survey of Income and Program Participation and Economic Data," by S. HABER, George Washington University.

8503 - "Using the Survey of Income and Program Participation for Research on the Older Population," by D.B. MCMILLEN, C.M. TAEUBER, and J. MARKS, Census Bureau.

8504 - "Summary of the Content of the 1984 Panel of the Survey of Income and Program Participation," by D.T. FRANKEL, Census Bureau.

8505 - "Enhancing Data From the Survey of Income and Program Participation With Data From Economic Censuses and Surveys," by D.K. SATER, Census Bureau.

8506 - "Methodologies for Imputing Longitudinal Survey Items," by V.J. HUGGINS, L. WEIDMAN, and M.E. SAMUHEL, Census Bureau.

8507 - "New Household Survey and the CPS: A Look at Labor Force Differences," by P.M. RYSCAVAGE, Census Bureau, and J. E. BREGGER, Bureau of Labor Statistics.

SIPP WORKING PAPERS - 1986

8601 - "Some Aspects of SIPP," compiled and edited by R.A. HERRIOT and D. KASPRZYK, Census Bureau.

8602 - "Nonsampling Error Issues in the SIPP," by G. KALTON, University of Michigan, and D.B. MCMILLEN and D. KASPRZYK, Census Bureau.

8603 - "An Investigation of Model-Based Imputation Procedures Using Data From the Income Survey Development Program," by V.J. HUGGINS and L. WEIDMAN, Census Bureau.

8604 - "Food Stamp Participation: A Comparison of SIPP With Administrative Records," by S. CARLSON and R. DALRYMPLE, Food and Nutrition Service.

8605 - "SIPP Longitudinal Household Estimation for the Proposed Longitudinal Definition," by L.R. ERNST, Census Bureau.

8606 - "A Comparison of Seven Imputation Procedures for the 1979 Panel of the Income Survey Development Program," by V.J. HUGGINS, Census Bureau.

8607 - "An Investigation of the Imputation of Monthly Earnings for the Survey of Income and Program Participation Using Regression Models," by V.J. HUGGINS and L. WEIDMAN, Census Bureau.

8608 - "Evaluation of Training Materials and Methods for the Survey of Income and Program Participation," by M. HOLT, Survey Research Consultant

8609 - "Patterns of Household Composition and Family Status Change," by C.F. CITRO, ASA/Census Research Fellow, and H.W. WATTS, Department of Economics, Columbia University

8610 - "Composite Estimation for SIPP: A Preliminary Report," by R.P. CHAKRABARTY, Census Bureau.

8611 - "Longitudinal Household Concepts in SIPP: Preliminary Results," by C.F. CITRO, ASA/Census Research Fellow, D.J. HERNANDEZ, and R.A. HERRIOT, Census Bureau.

8612 - "Following Children in the Survey of Income and Program Participation," by E.K. MCARTHUR, K.S. SHORT, and S. BIANCHI, Census Bureau.

8613 - "SIPP Labor Force Transitions: Problems and Promises," by P. RYSCAVAGE and K.S. SHORT, Census Bureau.

8614 - "Augmenting Data Reported in the Survey of Income and Program Participation With Administrative Record Data--A Brief Discussion," by D.K. SATER, Census Bureau.

Attachment 3

SIPP WORKING PAPERS - 1987

8701 - "Tracking Persons Over Time," by A.C. JEAN and E.K. MCARTHUR, Census Bureau

8702 - "Preliminary Data From the SIPP 1983-84 Longitudinal Research File," by J.F. CODER, D. BURKHEAD, A. FELDMAN-HARKINS, and J. MCNEIL, Census Bureau

8703 - "Work Experience Data From SIPP," by P. RYSCAVAGE and A. FELDMAN-HARKINS, Census Bureau

8704 - "The Treatment of Person-Wave Nonresponse in Longitudinal Surveys," by G. KALTON, J. LEPKOWSKI, S. HEERINGA, TING-KWONG LIN, and M.E. MILLER, Survey Research Center, University of Michigan

8705 - "SIPP: Filling Data Gaps on the Poverty and Social Welfare Fronts," by P. RYSCAVAGE, Census Bureau

8706 - "Response Errors in Labor Surveys: Comparisons of Self and Proxy," by D. HILL, University of Michigan

8707- "Differences Between SIPP and Food and Nutrition Service Program Data on Child Nutrition and WIC Program Participation," by L. KU and R. DALRYMPLE, Food and Nutrition Service, U.S. Department of Agriculture

8708 - "Quality Profile for the Survey of Income and Program Participation," by K. KING, R. PETRONI, and R. SINGH, Census Bureau

8709 - "Survey of Income and Program Participation (SIPP) Sample Loss and the Efforts to Reduce It," by D. NELSON, C. BOWIE, and A. WALKER, Census Bureau

8710 - "The Impact of Imputation Procedures on Distributional Characteristics of the Low Income Population," by P. DOYLE, Mathematica Policy Research, Inc., and R. DALRYMPLE, Food and Nutrition Service, U.S. Department of Agriculture

8711 - "Job Tenure, Lifetime Work Interruptions and Wage Differentials," by J. MCNEIL, E. LAMAS, Census Bureau, and S. HABER, George Washington University

8712 - "Measuring the Bias in Gross Flows in the Presence of Auto-Correlated Response Errors," by D. HUBBLE, Census Bureau, and D. JUDKINS, Westat, Inc.

8713 - "Investigation of Possible Causes of Transition Patterns From SIPP," by L. WEIDMAN, Census Bureau

8714 - "Households and Income Sources: Monthly Averages for 1984," by J. MOORMAN, Census Bureau

Attachment 4
COMPILATION OF ASA PAPERS
FOR 1984

PAPERS PRESENTED AT THE MEETING OF THE AMERICAN STATISTICAL ASSOCIATION: 1984
"SURVEY OF INCOME AND PROGRAM PARTICIPATION AND RELATED LONGITUDINAL SURVEYS: 198

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION I

"Analysis of Intra-Year Income Flows on the ISDP," by P. DOYLE, Mathematica Policy Research, Inc.

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

"The Measurement of Household Wealth in SIPP," by E.J. LAMAS and J.M. McNEIL, Bureau of the Census

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

"Using Selective Assessments of Income to Estimate Family Equivalences Scales: A Report on Work in Progress," by D. VAUGHAN, Social Security Administration

Discussant - COURTENAY SLATER, CEC Associates

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION II

"Toward a Longitudinal Definition of Households," by D.B. McMILLEN and R. HERRIOT, Bureau of the Census

"Lifetime Work Experience and Its Effects on Earnings: Data From the ISDP," by J.M. McNEIL, Bureau of the Census, and J.T. SALVO, New York City Department of Planning

"Panel Surveys as a Source of Migration Data," by D. DAHMANN, Bureau of the Census

"SIPP and CPS Labor Force Concepts: A Comparison," by P.M. RYSCAVAGE, Bureau of the Census

"Matching Economic Data to the Survey of Income and Program Participation: A Pilot Study," by S. HABER, George Washington University, and P.M. RYSCAVAGE, D. SATER, and V. VALDISERA, Bureau of the Census

Discussants - MARTIN DAVID, University of Wisconsin
HAROLD WATTS, Columbia University

PAPERS PRESENTED AT THE MEETING OF THE AMERICAN STATISTICAL ASSOCIATION: 1984

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION III

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

"Weighting of Persons for SIPP Longitudinal Tabulations," by L. ERNST, D. HUBBLE, D. JUDKINS, D.B. McMILLEN, and R. SINGH, Bureau of the Census

"Longitudinal Family and Household Estimation in SIPP," by L. ERNST, D. HUBBLE and D. JUDKINS, Bureau of the Census

"Longitudinal Item Imputation in a Complex Survey," by M.E. SAMUHEL and V. HUGGINS, Bureau of the Census

"Early Indications of Item Nonresponse on SIPP," by J. CODER and A. FELDMAN, Bureau of the Census

Discussant - ROY WHITMORE, Research Triangle Institute.

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION IV

"Month-to-Month Recipiency Turnover in the ISDP," by J.C. MOORE and D. KASPRZYK, Bureau of the Census

"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.

Discussants - GREG DUNCAN, Survey Research Center, University of Michigan
RICHARD ROCKWELL, Social Science Research Council

PAPERS PRESENTED AT THE MEETING OF THE AMERICAN STATISTICAL ASSOCIATION: 1984

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION V

"The Survey of Income and Program Participation," by R.A. HERRIOT and D. KASPRZYK, Bureau of the Census

"The German Socio-Economic Panel," by U. HANEFELD, Deutches Institute fur Wirtschaftsforschung, Federal Republic of Germany

"Household Market and Nonmarket Activities--The First Year of a Swedish Panel Study," by N.A. KLEVMARKEN, University of Goteborg, Sweden

"The Australian National Longitudinal Survey," by I. McRAE, Bureau of Labour Market Research, Australia

Attachment 5
COMPILATION OF ASA PAPERS
FOR 1985

PAPERS PRESENTED AT THE MEETING OF THE AMERICAN STATISTICAL ASSOCIATION: 1985
"SURVEY OF INCOME AND PROGRAM PARTICIPATION: 1985"

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION I

"Item Nonresponse in the Survey of Income and Program Participation," by D.B. MCMILLEN and D. KASPRZYK, Bureau of the Census

"Characteristics of Sample Attrition in the Survey of Income and Program Participation," by E.K. MCARTHUR and K.S. SHORT, Bureau of the Census

"Compensating for Wave Nonresponse in the 1979 ISDP Research Panel," by G. KALTON, J. LEPKOWSKI, and T.K. LIN, University of Michigan

"Alternative Definitions of Longitudinal Households in the Income Survey Development Program: Implications for Annual Statistics," by C.F. CITRO, Bureau of the Census

"Do We Learn From Past Experience When Constructing Complex Data?," by P. DOYLE, Mathematica Policy Research, Inc.

Discussants - JOHN CZAJKA, Mathematica Policy Research, Inc.
ROBERT TEITEL, Teitel Data Systems

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION II

"Operationalizing the Poverty Line in a Survey Using Subannual Reference Periods," by D. VAUGHAN, Social Security Administration

"Short Term Change in Household and Family Structure," by H. KOO, Research Triangle Institute

"Characteristics of Program Participants: Some Early Evidence From SIPP," by D.B. MCMILLEN, Bureau of the Census

"An Exploration of the Applicability of Hazards Models in Analyzing the Survey of Income and Program Participation: Labor Force Transitions," by K.S. SHORT and K.A. WOODROW, Bureau of the Census

"Gross Changes in Income Recipiency From the Survey of Income and Program Participation," by D. BURKHEAD and J. CODER, Bureau of the Census

"Designing a Data Center for SIPP: An Observatory for the Social Sciences," by M. DAVID, University of Wisconsin

Attachment 6
COMPILATION OF ASA PAPERS
FOR 1986

PAPERS PRESENTED AT THE MEETING OF THE AMERICAN STATISTICAL ASSOCIATION: 1986
"SURVEY OF INCOME AND PROGRAM PARTICIPATION: 1986"

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION I

"A Review of Research and Design Issues in the Survey of Income and Program Participation," by D. KASPRZYK, Bureau of the Census

"Effects of Adjustments for Wave Nonresponse on Panel Survey Estimates," by G. KALTON and M.E. MILLER, University of Michigan

"Life Events and Sample Attrition in the Survey of Income and Program Participation," by K.S. SHORT and E.K. MCARTHUR, Bureau of the Census

"Longitudinal Imputation for the SIPP," by S.G. HEERINGA and J.M. LEPKOWSKI, University of Michigan

Discussants - ROBERT MICHAEL, NORC
JOHN CZAJKA, Mathematica Policy Research, Inc.

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION II

"SIPP: Longitudinal Estimation for Persons' Characteristics," by E.L. KOBILARCIK and R.P. SINGH, Bureau of the Census

"Labor Force Transitions: A Comparison of Unemployment Estimates From Two Longitudinal Surveys," by M.S. HILL and D.H. HILL, University of Michigan

"An Additive Model of Recall Error: Analysis of SIPP Data," by D.H. HILL, University of Michigan

"Investigation of Gross Changes in Income Recipiency From the Survey of Income and Program Participation," by L. WEIDMAN, Bureau of the Census

"Measuring the Bias in Gross Flows in the Present of Auto-Correlated Response Errors," by D.L. HUBBLE and D.R. JUDKINS, Bureau of the Census

Discussant - MARTIN H. DAVID, University of Wisconsin-Madison

SURVEY OF INCOME AND PROGRAM PARTICIPATION: SESSION III

"Measuring Labor Market-Related Hardship Using SIPP Data," by B.W. KLEIN, Bureau of Labor Statistics

"Functional Limitations, Disabilities, and the Need for Personal Assistance: Data from the 1984 SIPP," by J.M. MCNEIL and C.J. HARPINE, Bureau of the Census

"Labor Mobility and Wages: New Data from SIPP," by P. RYSCAVAGE, Bureau of the Census

"Lifetime Labor Force Attachment: Retrospective Data From the Survey of Income and Program Participation," by E.J. LAMAS, J.M. MCNEIL, Bureau of the Census, and S.E. HABER, George Washington University

"Log-Linear Analyses for the Income Stability of Households in the Survey of Income and Program Participation," by H.A. SCHWARTZ, Bureau of the Census

"Longitudinal Household Concepts in SIPP," by CONSTANCE F. CITRO, National Academy of Sciences, D.J. HERNANDEZ and J.E. MOORMAN, Bureau of the Census

SURVEY OF INCOME AND PROGRAM PARTICIPATION AS A DATA BASE FOR PUBLIC POLICY

"Poverty Rates and Program Participation in the SIPP and the CPS," by R. WILLIAMS, Congressional Budget Office

"Factors Affecting the Earnings and Welfare Income of Unmarried Mothers," by T.N. GABE, J.E. GRIFFITH, and R.V. RIMKUNAS, Congressional Research Service

Discussants - JOHN FITZGERALD, Bowdoin College
EUGENE SMOLENSKY, University of Wisconsin-Madison

Attachment 7

JOURNAL OF ECONOMIC AND SOCIAL MEASUREMENT

Volume 13, Numbers 3 and 4, 1985

CONTENTS

Special Issue: Survey of Income and Program Participation

Guest Editor: Martin David

MARTIN DAVID
Introduction: The Design and Development of SIPP 215

Demographic Dynamics

HAROLD W. WATTS
The Scientific Potential of SIPP for Analysis of Living Arrangements for Families and Households 225

THOMAS J. ESPENSHADE and DOUGLAS A. WOLF
SIPP Data on Marriage, Separation, Divorce, and Remarriage: Problems, Opportunities, and Recommendations 229

GREG J. DUNCAN
A Framework for Tracking Family Relationships Over Time 237

Major Social Issues

REYNOLDS FARLEY
Understanding Racial Differences and Trends: How SIPP Can Assist 245

CAROLYN SHAW BELL
SIPP and the Female Condition 263

MARY JO BANE and JAMES WELSH
SIPP's Potential Contributions to Policy Research on Children 273

Enhancing Data on Human Capital

GARY S. FIELDS and GEORGE H. JAKUBSON
Labor Market Analysis Using SIPP 281

TIMOTHY M. SMEEDING
The Scientific Potential of SIPP: Its Content and Methods Regarding Fringe Benefits, Noncash Income, and the Value of Government Services 287

GAIL R. WILENSKY
SIPP and Health Care Issues 295

MICHAEL R. OLNECK
Critique of Questions Pertaining to Education in SIPP 299

Improving the Design of SIPP

MARTIN DAVID
The Distribution of Income in the United States Implications for the Design of the SIPP Panel 305

GRAHAM KALTON and JAMES LEPKOWSKI
Following Rules in SIPP 319

SHELDON E. HABER
A Perspective on Linking SIPP to Administrative and Statistical Records 331

JAMES D. SMITH
A Little SIPP: Old Wine in New Bottles--Let's Recask It 341

Analysis of Families Over Time

DAVID BYRON McMILLEN and ROGER HERRIOT
Toward a Longitudinal Definition of Households 349

GREG J. DUNCAN and MARTHA S. HILL
Conceptions of Longitudinal Households: Fertile or Futile? 361

MARTIN DAVID, RICHARD C. ROCKWELL, ALICE ROBBIN, and FRANKLIN W. MONFORT
Summary of the SIPP Conference and Recommendations of the Conference 377

MARTIN DAVID, ALICE ROBBIN, and RICHARD C. ROCKWELL
Summary and Recommendations: Second SSRC Symposium on the Scientific and Research Potential of SIPP--June 28-29, 1985 385

FOOD STAMP RESEARCH: RESULTS FROM THE INCOME SURVEY DEVELOPMENT PROGRAM AND THE PROMISE OF THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Table of Contents

I. OVERVIEW

A. Promise of SIPP
B. Background
C. Objectives of this Monograph
D. Organization of this Monograph

II. THE SIPP AND ISDP SURVEYS

A. The History of SIPP
B. The 1979 ISDP Research Panel
C. The Structure and Contents of SIPP

III. ECONOMIC STATUS OF THE FOOD STAMP TARGET POPULATION

A. Asset Holdings of Food Stamp Participant and Non-participant Households
B. The Pattern and Adequacy of Multiple Program Benefits
C. Multiple Program Benefits During the Course of a Year

IV. THE DYNAMICS OF FOOD STAMP RECIPIENCY

V. MONTHLY PATTERNS OF INCOME RECEIPT

VI. PROGRAM PARTICIPATION AND LABOR SUPPLY BEHAVIOR OF FOOD STAMP HOUSEHOLDS

A. Determinants of Participation in the Food Stamp Program
B. The Effects of Food Stamp Participation on the Market Labor of Female Heads of Household

VII. ISSUES IN THE USE OF PANEL DATA

A. Analytical Issues
B. Data Issues

VIII. ISDP RESEARCH AND THE PROMISE OF SIPP

REFERENCES

SIPP PUBLIC-USE FILES

At this time, tapes cost $175.00 each; so, for example, two tapes would cost $350.00. Technical documentation for the Relational (Complex) and Rectangular core files costs $15.00; for the Rectangular Topical file, $25.00.

(IMPORTANT: When ordering the Complex file from Customer Services, use the term ("Relational.")

Name of Data File	Number of Tapes 6250 bpi	1600 bpi
1984 Panel		
October 1983-January 1984:		
WAVE 1 RELATIONAL-CORE	3	11
WAVE 1 RECTANGULAR-CORE	2	8
February 1984-April 1984:		
WAVE 2 RELATIONAL-CORE	3	9
WAVE 2 RECTANGULAR-CORE	2	6
May 1984-August 1984:		
WAVE 3 RELATIONAL-CORE	3	9
WAVE 3 RECTANGULAR-CORE	2	8
WAVE 3 RECTANGULAR-TOPICAL	3	9
September 1984-December 1984:		
WAVE 4 RELATIONAL-CORE	3	12
WAVE 4 RECTANGULAR-CORE	2	8
WAVE 4 RECTANGULAR-TOPICAL	3	10
January 1985-April 1985:		
WAVE 5 RELATIONAL-CORE	3	12
WAVE 5 RECTANGULAR-CORE	2	7
WAVE 5 RECTANGULAR-TOPICAL	3	9
May 1985-August 1985:		
WAVE 6 RELATIONAL-CORE	3	10
WAVE 6 RECTANGULAR-CORE	2	7
*WAVE 6 RECTANGULAR-TOPICAL	NA	NA
September 1985-December 1985:		
WAVE 7 RELATIONAL-CORE	3	9
WAVE 7 RECTANGULAR-CORE	2	7
**WAVE 7 RECTANGULAR-TOPICAL	NA	NA
January 1986-April 1986:		
WAVE 8 RELATIONAL-CORE	3	7
WAVE 8 RECTANGULAR-CORE	2	5
**WAVE 8 RECTANGULAR-TOPICAL	NA	NA
May 1986-August 1986:		
WAVE 9 RELATIONAL-CORE	3	9
WAVE 9 RECTANGULAR-CORE	2	6
*WAVE 9 RECTANGULAR-TOPICAL	NA	NA
December 1983-March 1985:		
Annual Weights	1	1

*These files will be limited access releases. For further information concerning these files, contact Fred Cavanaugh, Population Division, 301-763-7207.

**Due to be released: Wave 7-September 1987, Wave 8-January 1988. For more current information, call Customer Services (301-763-4100).