

Daniel Kasprzyk and Charles A. Lininger, Social Security Administration

Introduction

The Survey of Income and Program Participation (SIPP) is a survey program proposed for implementation in 1982 after five years of development and testing by the Department of Health and Human Services in collaboration with the Bureau of the Census. The SIPP data base will contain information from periodic household interviews over 15-18 months matched to administrative record data, especially the Summary Earnings Record (SER) and Master Beneficiary Record (MBR) of Social Security benefits maintained by the Social Security Administration. A report on the activities of the development phase, the Income Survey Development Program (ISDP), and the goals and objectives of the SIPP are found in the 1980 Proceedings of the Survey Research Methods Section of the American Statistical Association ([6], [12]). Methodological analysis of experimental activities and evaluations of the quality of the data from the various collection activities of the ISDP are also found in the 1980 Proceedings ([1]-[5], [8]-[11]), and Reports from the Site Research Test ([7]). The SIPP has been specifically designed and proven in ISDP field tests to improve substantially the reporting of income and program-related data, and to provide the flexibility and timeliness to be responsive to the data requirements of program planning and policy analysis.

The SIPP is intended to provide:

- . monthly data on income, program participation, labor force activity, and household composition
- . assets and debts (of importance for asset-tested eligibility)
- . tax and disposable income data
- . measures of simultaneous receipt of benefits from cash and in-kind programs
- . routine matching of survey with administrative record data
- . new insights into income adequacy and poverty, based on improved measures of income, alternative reference periods, and the incorporation of assets and in-kind income
- . flexibility and responsiveness in adding special questions and in defining analysis units.

In general, the 1982 SIPP will provide an overview and more complete understanding of the tax and transfer system and its interrelation with the private sector.

The 1982 Survey of Income and Program Participation
Special Focus

The 1982 SIPP will also have a special analytical focus, the population aged 60-69. The last detailed survey of the aged (62+ years) was conducted in 1963; a short income survey of the aged (65+ years) was conducted in 1968. The SIPP will fill the need for data meeting Social Security and other departments' current policy research requirements with a supplemental sample of 5,000 households containing persons between the ages 60-69. The data will be used to provide a current data base for tabular and statistical analysis and microsimulation modelling. The 1982 SIPP is also intended to compliment results provided by another

study proposed to begin in 1982, of persons recently awarded social security benefits. The data will be useful in analyzing retirement at different ages, comparing labor force histories of social security beneficiaries and non-beneficiaries at different ages, and examining their relative age-specific health characteristics and ability to work.

Survey Design

Interviews with persons in the 1982 SIPP sample are to be conducted four times at four month intervals beginning in February 1982 and concluding in May 1983; an additional fifth interview, designed specifically to collect tax data, is possible. Each wave of interviewing will consist of four rotation groups, with approximately the same number of households assigned randomly to each of the four groups. The reference or recall period for most data items will be four months.

Census Bureau interviewers will collect the information during personal visits to the sampled households. Interviews will be obtained for each person aged 15 and over in the original sampled household. The first or bounding interview described below averages approximately 25 minutes for each person. Later interviews will be slightly shorter, about 20 minutes. Self responses will be obtained for all adults present at the time of the interview and proxy responses accepted for all others.

A basic feature of the survey design is that the panel sample is made up of persons (not addresses as in the CPS) who, if they move, are followed to their new addresses as long as the new address is within 50 miles of any SIPP Primary Sampling Unit (PSU). Anyone living with an original sample person, either at the sampled or new address, will also be interviewed in order to continue providing the social and economic context in which sample persons reside.

Sample Design

The 1982 SIPP sample will consist of 20,000 occupied living quarters (interviewed and noninterviewed households) in 164 Primary Sampling Units across the country. Approximately 15,000 of these will be a national probability sample of addresses drawn by the Census Bureau from the "1970 Census Current Surveys Design C Sample Reserve Measures," updated for new construction. Within each sample PSU, a systematic sample of clusters of addresses (expected cluster size is 4) is selected with the same overall selection probability. The sample drawn from this general population frame is, thus, self-weighting, since every sample unit will have the same probability of selection.

The remaining 5,000 occupied living quarters will be selected to increase the representation of the population aged 60-69. Two sampling frames will be used to oversample this subgroup:

- (1) The Health Insurance Master File administered by the Health Care Financing Administration, which contains all

persons enrolled in the Medicare program. Its coverage of the 65+ population is, therefore, reasonably complete (about 95%). Approximately 850 persons, enrolled in the program as of August 1981, will be included in the supplemental sample to increase representation of persons, aged 66-69 (as of December 31, 1982) who are not receiving Social Security or Railroad Retirement benefits.

- (2) The 1980 Census 100 percent Final Detail File of the Bureau of the Census. This file will be used to obtain approximately 4,150 occupied housing units containing a person between the ages 60-69 (as of December 31, 1982).

The proposed distribution of the supplemental sample is given in Table 1. The expected number of SIPP sample households with 60+ aged units from both area and supplemental samples is given in Table 2.

Survey Content

Questionnaires will vary from wave to wave in the SIPP Panel. There are two general content components to the questionnaires: first, the "core" portion which changes little, if at all, over the life of the panel; and second, the topical sections, which vary between waves to address timely policy issues with minimal lead time and to obtain significant information (such as assets, debts and categorical eligibility criteria) that is less volatile or more easily recalled over long intervals.

Core items deal with changing household composition, labor force status, and an income profile obtained in the first interview and updated in later ones. The initial interview (February 1982 - May 1982) is devoted mainly to measuring all sources of income for each person fifteen years of age and older in sample. Specific questions are asked about 1) labor force activity, 2) receipt of income associated with retirement, disability, unemployment, etc., 3) receipt and amounts of income from means tested government transfer programs, and 4) ownership of assets and income flows from them. When persons report self-employment income and/or owned a business or farm, there are also questions about the nature of the job, business or farm, and the amount of income received during the reference period. The receipt and amount of income going to children under the age of fifteen in the household is reported by a responsible adult.

This detailed income source profile is retained and used in the following waves to update income receipt and amounts, along with changes in household composition and labor force participation. The update procedure requires less time than the initial interview, allowing interviewers to ask additional special question sets ("topical modules") on new subjects of policy interest. Thus, the second SIPP interview will collect data needed to simulate programs in the federal transfer system -- assets, debts, disability status, work related expenses, child care and medical costs -- as well as determine eligibility status and payment amounts for a number of federal transfer programs.

In the third interview (October 1982 - January 1983) the income profile will be updated once again. The proposed topical focus of this interview will meet the needs of the Division of Retirement and Survivors Studies in the Office of Research and Statistics within the Social Security Administration. Data elements proposed for this wave include the need for care, informal assistance received and provided, transfers within and across households, marital history (linked when possible to SSA earnings records of past and present spouses to permit studies of earnings sharing proposals), work history, pension coverage and vested rights, and plans or past decisions to retire.

In the fourth interview (February 1983 - May 1983) the income profile of each person in sample is updated once again to complete the income profile for the full calendar year, 1982. The collection of details on taxes, filing status and claims for dependents during 1982 will begin in this interview wave with additional follow-up continuing beyond the end of the interviewing period as necessary. The remaining topical content of this interview has not yet been assigned and is available for priority issues.

Special Features of the 1982 SIPP

A number of features of the 1982 SIPP improve upon and extend the work of the development program's activities. For example:

1. Links to Administrative Records and the Validation of Social Security Numbers

From the survey's inception, The SIPP data system has always been planned as a combination of data from administrative records and household surveys. This reduces respondent burden by looking to other data sources for difficult-to-obtain information. Household interview responses will be supplemented by information from program files such as the earnings and benefit records of the Social Security Administration. This allows, for example, analysis of the long-term impact of various Social Security benefit formulas, which are a function of past earnings histories and retirement decisions of currently ineligible workers rather than the characteristics of current recipients. To make these linkages accurate, social security numbers are required for sample individuals. During the developmental program, social security numbers were obtained and verified for more than 95% of the sample; the 1982 SIPP is expected to equal or exceed this rate. At the conclusion of each month's interviewing during the first interview wave of the 1982 SIPP, a special extract file will be prepared by the Bureau of the Census for the Social Security Administration. This file will contain a small number of key variables in a format appropriate for machine validation of survey reported SSNs -- SSN, full name, date of birth, sex, race, and unique SIPP identification number. The Social Security Administration will then identify (by machine validation) incorrectly reported SSNs within one month of receipt so that "problem" cases

can be flagged and returned to the field prior to the second round of interviewing. During Wave 2 every household with one or more persons who did not report a valid SSN (as identified in the machine validation) will be asked to verify or correct the SSN and the critical matching variables obtained in Wave 1. At the conclusion of Wave 2, another file will be prepared for SSA in order to machine validate the problem cases. Cases which remain either missing or invalid will be clerically resolved.

Social Security Numbers of persons who enter the panel after the first wave (because they live in households with original sample people) will be machine validated at the conclusion of the Wave 4 interviews; SSNs not verified during this process will also be subject to clerical resolution. Finally, to improve SSN reporting, the SSN will be flagged as a critical item on the control card to emphasize its importance to interviewers.

These procedures should provide accurate and timely SSNs for over 95% of the sample, and make possible early links to SSA's Master Beneficiary Record (MBR) and Summary Earnings Record (SER).

2. Quality Control

All survey programs have their own quality control procedures to ensure an acceptable data set. Experience from the development program indicates that well defined procedures, above and beyond standard quality control, are necessary in a panel survey of SIPP's complexity. Thus, the Census Bureau is developing a management plan detailing all aspects of the survey program. The plan will identify the Census Bureau's responsibilities, detailed schedules and deadlines, an overview of procedures by task area, and quality control standards for major aspects of the field work. The ultimate focus of the quality control program will be on procedures which monitor response rates, attrition, and time lags in data entry and computer software development.

Specific programs in the field to ensure data quality include:

- a. A reinterview program to control the interviewers' performance. This will allow the Supervisory Field Representative to ascertain whether: (i) the interview was conducted; (ii) the control card data accurate, particularly household composition, (iii) noninterview types correctly identified; and (iv) original responses for parts of the questionnaire recorded correctly. Each month one-fourth of the interviewer assignments of one-fourth of the interviewers will be selected for the reinterview sample in Wave 1, so that every interviewer will be included in the reinterview program during Wave 1. During Waves 2 and 3, one-eighth of the interviewers and one-fourth of their assignments will be selected each month, so that by the conclusion of the third wave each

interviewer will have been included in the reinterview program twice.

- b. As part of the processing operations, each regional office will be supplied with computerized information which serves as quality control checks for both interviewers and regional office staff. Interviewer performance tables will be produced every month for each regional office. These tables will include for each interviewer pre-edit errors by error type and noninterview rates by type of noninterview. Thus, each interviewer's performance can be followed closely by regional office staff.

Strict interviewer performance standards will be set based on observation, error rates, production ratios, and noninterview rates. Monthly and quarterly performance reports as well as an official performance report at the end of each wave will be given to each interviewer.
- c. New clerks in the regional office will receive 100% verification of their editing and coding, while for new keyers there will be 100% verification until the number of errors is less than .5%.
- d. Field observations of each interviewer will be conducted by the SIPP supervisor or a senior interviewer after initial training.

The development program has also shown the need to tighten control on maintaining the sample. Thus, a longitudinal control file will be instituted to ensure that each person or household record entering the regional office pre-edit is the same person or household expected in the rotation and that the entire sample is accounted for in each wave. The control file will contain identifiers and demographic information necessary for matching persons between waves. The establishment of an accurate longitudinal control file will ensure routine and timely cross-wave linkages.

3. Regional Office Data Entry

The SIPP development program featured centralized processing control by the Census Bureau's Data Processing Division in Jeffersonville, Indiana. Receipts from the field were checked in and compared to a control list of expected returns, and notification of instrument nonreceipts was transmitted to the regional offices. Reconciliation and additional transmittals from the field offices took place at this stage. This process has obvious built-in delays, since it relies on mail and other communications between the centralized operation and the field offices. Thus, the correction of problem data by the field staff necessarily lagged far behind the original interview.

The 1982 SIPP will attempt to resolve these problems by using ENTREX equipment installed in the regional offices as part of the 1980 Census operations. This equipment will allow the regional office to pro-

cess questionnaires, including all required data keying activities. Edits of some fixed-field and source coded data -- such as source code sequence checking and range checking -- are also included as part of this system. Keyed questionnaire data will be transmitted to Washington at the end of each processing cycle, allowing timely questionnaire check-in, an extensive edit, and the establishment of a longitudinal control file for recording changes in family and household composition over time. Errors, missing data items, and nonreceipts will be quickly communicated to the appropriate regional office for reconciliation by staff through office review or interviewer follow-up.

4. Multiple Frame Estimation

The development program, because of its use of several sampling frames in each data collection, has shown that multiple frame weighted data cannot be made available soon after data collection without proper planning, staffing and attention to detail. Several steps are being taken to ensure the timely derivation of multiple frame weights.

A preliminary multiple frame weight will be obtained by using first-interview responses to identify whether a sample element had a chance of being selected into sample from more than one sampling frame. Thus, each person reporting enrollment in Medicare will be asked the effective date of coverage to identify whether she/he had been on the Medicare file at the time of sample selection. Fortunately, the effective date appears on the Medicare enrollment card, and Medicare enrollees are routinely asked for their cards during the interview in order to record the type of Medicare coverage.

Whether a sample person resided in the SIPP universe at the reference time of the 1980 Census File is slightly more difficult to determine. Each sample person aged 60-69 as of December 31, 1982 will be asked whether he/she was living outside the United States, or in a nursing home or institution on April 1, 1980. A classification of each sample household into its correct domain will be based on the responses to these questions, and a preliminary weight assigned to each case. In addition, codes appended to the sample master file will allow the computation of sampling error estimates quickly and speed the process of obtaining the coefficients of the multiple frame estimators since these estimators are functions of variances and covariances.

5. Timely Delivery of Data Products

Both HHS and Census are committed to an early delivery of data products to analysts. This commitment requires early completion of questionnaire drafts in order to have adequate time to develop and test computer software before the actual data are collected. Over 90% of the specific data items will be final 9-10 months prior to interviewing,

leaving up to 10% of the questionnaire available for data items inserted to respond to last minute policy initiatives. At the time of writing, the questionnaires for the first two waves had been pre-tested, the first was in final form, and the processing system was being developed. Early agreement on questionnaire content for later interviews should help in obtaining data relatively soon after data collection.

Timely availability of survey data also depends on an early agreement between the sponsors as to what constitutes the survey data products. Although HHS and Census requirements diverge somewhat, discussions are continuing to resolve differences and ensure timely delivery. HHS has requested a number of preliminary files -- pre-edited and consistency edited cross-sectional files, which resemble the questionnaire. There are, however, only five primary survey data products emerging from the 1982 SIPP.

The first will be a Wave 1 cross-sectional analysis file containing less detail than the questionnaire, summary income variables for persons, families, and households, and imputations for income and income-related item nonresponse. This file will also be weighted to population controls by age, race, and sex; it will have a single frame weight as well as a multiple frame weight. The latter will be a preliminary weight based on survey responses. The determination of a final weight based on identifying multiple chances for selection by matching the survey respondents to the sampling universe files is under study.

The second primary product will be a Wave 1- Wave 2 cross-wave analysis file. Assurances by Census that the linking identification numbers will be accurate, and sample control will be improved suggest that persons and households from the first two waves of SIPP should be linked quickly and accurately. This linked file will contain cross-wave edits and imputations of income and work experience, and cross-wave edits of selected items ("program eligibility" items) in the topical section of the questionnaire. Codes to identify family and household membership for each month will also be created.

The third primary product will be the topical module data from the third and fourth waves of interviewing. SIPP identification numbers and linking variables will allow the topical modules to be merged easily and quickly with other SIPP data products -- the Wave 1 - Wave 2 cross-wave analysis file, for example.

The fourth primary product will be the SIPP Panel Master Longitudinal File, a concatenation of all data collected in the survey. The critical aspects of this file are that a) persons, households and job levels are matched accurately, and b) personal identifiers are correct and allow accurate reconstruction of families and households at any time during the reference year. Some limited longitudinal editing and imputation may be implemented on this file.

Finally, the fifth primary product will be an annual analysis file which summarizes data for the 1982 calendar year. The file will contain a) historical and control data on households, families and persons, b) monthly income and summary income variables by type, c) annual income variables for persons derived from the monthly income variables, d) cross-wave edits and imputations of income and work experience, e) tax information, f) codes to identify monthly family and household membership, and g) subsets of topical module data, primarily those data used in microsimulation modelling. As mentioned above, early deadlines for questionnaire development and early specifications of the details of the files should ensure timely availability of the survey data. Other files, such as public use files and survey files linked to administrative data, will also be produced. However, most data requirements for policy analysis should be satisfied by the data products cited above.

SIPP Research and Analysis in Progress

While the focus of this paper has been the 1982 SIPP Panel, the HHS-SIPP Staff's responsibilities extend beyond the survey's development and implementation. Preliminary substantive and methodological results from the development program are now appearing in increasing numbers. HHS-SIPP staff projects include 1) the analysis of multiple receipt of benefits in five programs (Social Security, Supplemental Security Income, food stamps, public assistance, and Unemployment Compensation), 2) the analysis of multiple receipt of benefits in means tested programs -- cash and in-kind program benefits, 3) a description of the assets of food stamp participants, 4) an analysis of the receipt of multiple pension benefits by OASDI disability beneficiaries, 5) a description of the characteristics of persons in poverty, (6) an overview of the economic status of the aged, including income, assets, and household characteristics, and 7) a description of the benefits received through the emergency heating assistance program during Winter 1979-1980. While primarily descriptive and preliminary, these analyses use data from the 1979 ISDP Panel which have never before been available. Analyses such as these preview the types of analyses which can and will be done using the larger, less experimental 1982 SIPP.

Methodological analysis has been an important part of the development program; it will continue to be important in the ongoing SIPP program. The development program established four multi-year research centers in the areas of (1) nonresponse and imputation, (2) microsimulation modelling, (3) net worth, and (4) income. Each center not only responds to the analytic initiatives of the SIPP staff, but also identifies and proposes promising research and applications of the ISDP-SIPP data in its area of expertise. This work is helping the SIPP staff to use the 1979 Research Panel data and improve their planning for the 1982 SIPP. The work done under contract to these centers has been both specific to the 1979 Panel and general to survey research issues. A complete description may be found in the forthcoming SIPP Research Note,

Survey Development Research Centers. The SIPP program continues to expect to fund research in broad areas of survey and statistical methodology using both the development program data and the SIPP data. The remaining papers of this session are the result of work currently being done by the ISDP-SIPP Nonresponse and Imputation Research Center at the University of Michigan.

Funding

At the time of the ASA meeting in August 1981, the funding situation for the 1982 SIPP Panel was ambiguous. Eight million dollars and twenty-six positions were included in the Administration's Budget before Congress, but the Department of Health and Human Services was reconsidering its priorities. Subsequently, the Secretary decided to withdraw support, and work relating to the 1982 SIPP Panel has now been terminated.

The Income Survey Development Program is, however, continuing; the HHS staff are engaged in evaluation and analysis of the ISDP data and in the preparation of survey documentation and public use tapes from the 1979 ISDP Research Panel.

REFERENCES

- [1] Coder, J. F. (1980). Some Results from the 1979 Income Survey Development Program Research Panel. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [2] Feldman, A., Nelson, C., and Coder, J. F. (1980). Evaluation of Wage and Salary Income Reporting on the 1978 Income Survey Development Program Test Panel. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [3] Kalton, G., Kasprzyk, D., and Santos, R. (1980). Some Problems of Nonresponse and Nonresponse Adjustment in the Survey of Income and Program Participation. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [4] Klein, B. W. (1980). How the SIPP can Respond to the Recommendations of the Levitan Commission. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [5] Lane, J. A., and Olson, J. A. (1980). 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.
- [6] Lininger, C. A. (1980). The Goals and Objectives of the Survey of Income and Program Participation. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

- [7] Olson, J. A. (Editor), (1980). Reports from the Site Research Test. Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services.
- [8] Olson, J. A., and Klein, R. E. (1980). Interviewers Perceptions of Reasons for Participation Refusal in a National Longitudinal Survey, 1979 - 1980. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [9] Vaughan, D. R., and Lancaster, C. G. (1980). 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.
- [10] Wertheimer, R. F. (1980). The Role of SIPP in Microsimulation. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [11] Ycas, M., and Kasprzyk, D. (1980). Operationalizing Alternative Disability Definitions in a Household Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.
- [12] Ycas, M., and Lininger, C. A. (1980). The Income Survey Development Program: A Review. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1980.

Table 1

Proposed Distribution of the Supplemental Sample of Persons Aged 60-69 ^{1/}

	<u>1980 Census Sample</u>				<u>Medicare Sample</u>	<u>Supplemental Sample</u>
	<u>60-61</u>	<u>62-64</u>	<u>65-69</u>	<u>Subtotal</u>	<u>66-69</u>	<u>Total</u>
OASDI Beneficiaries	40	980	1,320	2,340	-	2,340
Non-beneficiaries	385	1170	255	1,810	850	2,660
Total	425	2,150	1,575	4,150	850	5,000

^{1/} Beneficiary/non-beneficiary dichotomy in the Census sample is an estimate based on population distribution.

Table 2

Expected Distribution of SIPP 60+ Aged Units: Area Plus Supplemental Sample ^{2/}

	<u>Age</u>				
	<u>60-61</u>	<u>62-64</u>	<u>65-69</u>	<u>70+</u>	<u>Total</u>
OASDI Beneficiaries	85	1,357	2,112	1,828	5,382
Non-beneficiaries	885	1,612	1,305	377	4,179
Total	970	2,969	3,417	2,205	9,561

^{2/} Aged units are defined as (1) a married couple living together with one or more members 60 or older, or (2) an individual 60 or older who was widowed, divorced, living apart from his/her spouse, or never married.