Reengineering the SIPP: The New Dynamics of Economic Well-being System
Brookings/Census Roundtable
June 8, 2006

INTRODUCTION

The comprehensive information about individual and household income and program participation collected by the Survey of Income and Program Participation (SIPP) is used by federal agencies to evaluate programs and assess need. A major use of the SIPP has been to evaluate the effectiveness of government programs and to analyze the impacts of options for modifying them. The Social Security Administration, for example, relies on SIPP data to project baby boomers' retirement incomes and the likely timing of their retirement. The Department of Health and Human Services uses SIPP data to evaluate the impact of Welfare Reform and to measure the economic effect of disabling conditions on children and adults.

The SIPP's longitudinal design has many advantages, but imposes considerable burden on respondents and makes review and data processing difficult and time consuming. The re-engineered system, to be known as the dynamics of economic well-being system, is expected to reduce respondent burden and attrition and deliver data on a timely basis. Although it will not supply the same level of detail as the SIPP, its design must offer policymakers and researchers data that address the same basic issues. Several options are now being considered for a new system that will provide information on measuring the dynamics of economic well-being.

The dynamics of economic well-being system will take advantage of the advances that the Census Bureau has made in acquiring and integrating administrative records with survey data, in modeling local area estimates, as demonstrated by the current Small Area Income and Poverty Estimates (SAIPE) and Small Area Health Insurance Estimates (SAHIE) programs, and in developing new data collection systems like the American Community Survey (ACS). The new system will use data from current demographic surveys and administrative records to identify a population cohort that will be measured longitudinally by using both administrative data combined with a new demographic survey instrument. A major goal of this new system is to develop monthly estimates of whether and how much individuals participate in cash assistance programs, and to include a longitudinal component.

SYSTEM OBJECTIVE AND SCOPE

The overall objective of the dynamics of economic well-being system is to reengineer the current SIPP to construct a streamlined system that can provide similar information at a reduced cost. The system will be able to generate data that can be used, in part, as SIPP data have been used.

The scope of this plan includes those activities that must be undertaken in order to implement a prototype for use in FY 2008. The first requirement is to identify core elements to be included in the new data system. In addition to identifying core data elements to be included, several configurations of the data collection system are to be examined. Once a prototype has been implemented, requisite changes can be made and included in a subsequent plan. The final prototype evolves as design elements are tested and evaluated.

The first data product from the new dynamics of economic well-being system is planned for release to the data user community in December 2008. This file will be a two-year retrospective file of a sample of respondents based on an existing demographic survey enhanced with current administrative records data.

The two-year retrospective file from the new system will consist of data from the common sample from the selected current survey for calendar years 2006 and 2007 plus administrative records. Therefore the scope of this data will not be as complete as the current SIPP data. The file of data from the same sample using the new survey instrument the following year will provide both annual and sub-annual data for 2008.

COMMUNICATIONS PLAN

Policy makers, including several major federal agencies, need detailed information on the dynamics of income, poverty, wealth, health insurance coverage, program participation and other aspects of economic well-being. SIPP has provided this information for the past two decades. No other statistical program provides the necessary level of detail. The Census Bureau’s plans for a re-engineered system to measure economic well-being will continue to meet the most important of these information needs, in a timelier manner at reduced cost. To ensure the satisfaction of SIPP stakeholders, they will continue to be consulted on decisions throughout the development process.

Census management will continuously meet with key stakeholders to assess their priorities for the new system. The Census Bureau has presented overviews of the new system to various audiences, including Council of Professional Associations on Federal Statistics (COPAFS), Census Advisory Committee of Professional Associations, the Congressional Budget Office, Congressional committees, Congressional hearings, and key federal stakeholders. Census staff will organize regular meetings with federal staff and key outside researchers to assess their needs and apprise them of the progress of the new system.

SYSTEM OBJECTIVES

The *Program Development Team* was established to oversee the new dynamics of economic well-being system. The team defined 5 major objectives that are integral to the design, development, and implementation of the new system. An inter-divisional project group was established to oversee each objective. Below is the progress to date for each group.

I. PROTOTYPE DEVELOPMENT

The *Prototype Development Group* is responsible for all phases of the design and development of an integrated current survey and administrative records-based prototype system and resulting data set. Prototype development is a continuing process that will evolve as the entire system evolves. The early prototypes will be used mainly to demonstrate 'proof of concept' and determine disclosure issues associated with creating a Public Use File containing administrative records data. Of particular importance in this activity is to study both the use of administrative records as a measure of data quality, for use in edits and imputations, and as replacement for respondent's data. First, however, the quality of the administrative data themselves must be addressed. Later prototypes will more closely reflect the final product. At every phase of development, the resulting data file will be reviewed by the Census Bureau's Disclosure Review Board.

The first prototype is a combination of one year of data from the March, 2001 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) and Medicaid data from the 2001 MSIS. The second prototype, which is currently being developed, will consist of two years of ASEC data, Medicare data[1], and perhaps synthetic data for asset income.

II. ADMINISTRATIVE TO SIPP ASSESSMENT

The *Administrative Records to SIPP Assessment Group* is involved in ongoing evaluations to determine the current availability of administrative data sources that have sufficient national coverage and can be obtained in a consistent and timely manner. These evaluations are iterative, and will continue as the design progresses and the new survey content becomes more solidified. As a first step the group attempted to align core variables from 1993 SIPP Panel longitudinal file with currently available administrative records data including identifying the variable(s), defining the coverage of the data source, and determining the lag between when the administrative data are collected and when they are available at the Census Bureau. Typical match rates for administrative data to survey data are in the low 90 percent range. Examples, by core section, are discussed below.

---

[1] Medicaid data was dropped from the second prototype because the time lag for the MSIS is too long.

Health Insurance

- ***Medicare*** The majority of SIPP items align with a current administrative records source and a national-level file is available in April of each calendar year.

- ***Medicaid*** SIPP Medicaid items align with a current national-level administrative records source. However, there is a 3-year lag in obtaining this data. There are no administrative data sources for Medicaid state expansion or other types of public health insurance.

Assets

- ***Asset ownership, gross income, and net income*** All aggregate asset SIPP items (with the exception of ownership of and income from mortgages) align with current Internal Revenue Service (IRS) administrative records sources—IRS 1040, 1099-INT, and 1099-DIV. These are national-level files have negligible delays in delivery to Census.

Labor Force

- ***Total earnings*** Total earnings from a job(s) or income (loss) from a business(es) align with administrative data available from the IRS 1040. The New Survey Group is currently evaluating an event history calendar which would be used to collect data about labor force experiences.

Demographics

- ***Demographic characteristics*** Data from the Census Numident and the Person Characteristics File (PCS), both national files, are available for some respondent characteristics (e.g., age, Hispanic origin, race, sex, U.S. citizenship).

Many programs are administered at the state level. At the current time, there are limited state data to address receipt of these programs. However, this is an area that we will look at closely to determine which state partnerships can be developed in order to obtain the necessary data files in the future.

III. **SURVEY CONTENT, METHODOLOGY, AND DESIGN**

A. **Survey Content**

The *Content Group* will determine the variables measured in the new dynamics of economic well-being system survey instrument. The group is using the 1993 SIPP Panel longitudinal file as the starting point. Using this file as a starting point is advantageous because it limits the scope and the number of items to those previously required for longitudinal data analysis. Additional variables will be added and/or current variables eliminated based on the outcome of this and future stakeholder meetings and Census analytical requirements. Other factors, such as the current availability of administrative records, the budget, and the final survey methodology—including mode of data collection (which is discussed in the next section), will also impact the final survey content.

Because of changes in SIPP content since the 1993 longitudinal file, a brief synopsis of each section of the 1993 longitudinal file follows:

Labor force participation (e.g., current employment status, weeks worked, weeks spent looking for work or on layoff); ***receipt from general income sources*** (e.g., retirement, pension, disability, Social Security, Medicare, Supplemental Security Income (SSI)); ***public and private health insurance coverage***; ***asset ownership***; and, ***school enrollment and assistance***.

If ***employed by a job or jobs***—the number and name of each employer, type of business or industry, the type of work, the frequency and amount of pay. If ***owned a business or businesses***—the type of business and work, the number of hours worked per week, the total number of employees.

In addition to the above components, there is also an Amounts section and a Program section. The Amounts section determines the period of eligibility, the family members covered, and the dollar amounts received for each income source. The Programs section determines the household's participation in rent and/or energy assistance programs as well as the National School meals program.

1. <u>Decision Matrices–Content</u>

   A series of matrices have been developed to help organize discussion. Input was solicited from the three areas in order to derive the final survey content—they are: key stakeholders, data analysts, and administrative records staff. The goal of these matrices is to provide a complete picture of the user's needs and current availability of data.

Because of budget restrictions and concerns about respondent burden, our vision is that the content of the new survey will be to supplement that of the existing survey and administrative records data as well as serve as a screening device for future topic surveys. In other words, two or three questions about child well-being in the new survey could identify the universe of respondents for a follow-on survey on the well being of children. Once established, the core content of the new survey should remain constant.

B. **Survey Methodology and Design**

In conjunction with the Content Group, the *New Survey Group* is leading the development of the structure and components for a new data collection system. The group is developing potential options under the following assumptions:

- The system will collect data covering at least three years and have the ability to provide required sub-annual data.
- Some portion of the data will be obtained from administrative records.
- Timeliness of the data release with respect to data collection is crucial.
- The system will build upon data and data processing systems for current surveys, in particular the ACS and the CPS.
- Sampling, estimation, and disclosure-proofing strategies can be developed to support the options being considered.

1. <u>Level of Detail</u>

The historical precedent set by the SIPP program is that of detailed data on a core set of topics, with topical modules of a broad and varied scope added for each wave of interviewing. The scope and detail contributed to the complexity of the instrument, the length of the interview, and the high attrition rates, all of which we hope to reduce with the new survey. The key is that reducing the content detail will produce a more streamlined survey that will be less costly and reduce respondent burden.

2. <u>Periodicity and Question Format</u>

The SIPP program administered three waves or interviews in each field year. The interview rotation design meant that data were collected over all 12 months of the calendar year. This required a constant level of production-readiness which precluded the proper level of focus for either the preparatory or the post-collection processes, contributing to data products that were not timely and were error-prone.

Two alternative options are currently being evaluated. The first involves a single survey contact in the household with a reference period long enough to capture retrospective information on the required dynamics, but requires a non-traditional approach to obtain month-to-month transitions—such as an electronic event history calendar. The second option consists of an initial contact with follow-up interviews at 6-month intervals (the reference period would be the preceding 6 months). Considerations and research of option effects on seam bias, nonresponse, etc. must be considered when deciding among these options.

3. Mode

Traditionally, SIPP has relied on personal visits by field staff to establish and maintain rapport with the sample cohort, although over the life of a longitudinal panel, telephone interviews are encouraged in order to minimize costs.

Although it is widely assumed that interviewer-administered surveys are of higher quality, the potential exists to cut costs without significantly impacting data quality by implementing multiple collection modes such as Computer Assisted Telephone Interviewing (CATI), decentralized telephone interviewing, and mail out/mail back questionnaires as long as the questionnaire is not overly complex.[2]

The mode for the new data collection system is highly dependent on the complexity of the questionnaire, the budget, and the desired response rate.

4. Sample Source and Size

There are several options for generating a sample frame[3], each with different characteristics and advantages. These options include:

- sampling from the Master Address File (MAF) both with and without auxiliary data;
- sampling directly from ACS interviewed cases;
- sampling directly from administrative data;
- selecting all CPS cases matched from one ASEC to the following year's ASEC;
- using all CPS cases in outgoing rotations (all months); and
- using the existing SIPP sample[4] already identified and unduplicated.

[2] The American Community Survey (ACS) has had much success with their mailout-CATI-CAPI design, which is being considered as an option.

[3] These options will be discussed more fully in the sample design section of this document.

[4] The 2000 sample redesign selected enough SIPP cases to cover the 10-year period between decennial censuses. The SIPP sample cases that have not been sent to the field are available for the dynamics of economic well-being system.

The sample size is determined by the level of reliability required for specific estimates as well as the survey's budget. The sample size required to produce only reliable national estimates is much smaller than what would be needed to produce reliable state estimates.

As a source of efficiency and cost effectiveness, several options being considered utilize some existing and processed data collected such as the ACS, although the CPS is also under consideration. Both sources of data provide enough background household demographics, income, and labor force information to serve as characteristics and analytic stratum for examining program and other transition information collected by the new survey component of the dynamics of economic well-being system.

Starting with data collected from the ACS in any given year or specific months within the year, we can conduct another ACS interview the following year followed by the dynamics of economic well-being system questionnaire in year 3. Another option is to use a new questionnaire in both year 2 and year 3 (and later years, if funded by other sources.)

Another alternative is to use the CPS basic or ASEC data as the sample for the dynamics of economic well-being system interview. The CPS design allows for a 50 percent overlap from year to year and a 75 percent sample overlap from month to month. By utilizing this design characteristic, an existing longitudinal product and sample is available to interview for a third year, the year after they would have rotated out of sample in CPS. This third interview can use the new dynamics of economic well-being system questionnaire.

5. <u>Following Movers</u>

All of the options being discussed have some component where following movers is required. The simplest option involves identifying a sample from an existing survey source (such as the ACS) and then locating the sample one year later for an interview. The same instrument could be used in order to provide another point-in-time measure.

Several additional options for locating movers become available when administrative records from Social Security, the IRS, and the National Change of Address information are used.

6. <u>Survey Design</u>

The current SIPP design includes an oversample for low-income households. This may still be a requirement for the dynamics of economic well-being system, but could shift to oversampling of other subgroups (the elderly, families with children, or race/Hispanic subgroups). For any oversampled group, auxiliary data is needed to determine which cases are in the oversample group. If oversampling is based on a characteristic that changes over time, greater sampling efficiencies will be possible by interviewing within a short time frame following the sample selection. For example, sampling directly from ACS will likely result in the highest oversampling efficiency, while sampling from the MAF with auxiliary data attached will be a little less efficient.

The current SIPP design is clustered into Primary Sampling Units (PSUs) for efficiencies in field representative workloads. In order to continue with this PSU type design, the sampling frame will need to contain a sufficient number of cases within those sampled areas in order to ensure equitably sized workloads.

Finally, the coverage of the frame will need to be taken into account in evaluating the alternative sources for the sample.

IV. **DATA PRODUCTS**

The main product of the new dynamics of economic well-being system is an ongoing longitudinal micro-data file that is publicly released in a timely manner. The first data product from the dynamics of economic well-being system is planned for release to the data user community in December 2008. The resulting file will be a two-year retrospective based on an existing demographic survey enhanced with current administrative records data. The third year of the system, in the form of the new survey instrument, will be fielded in the following year. One alternative is to have all three years of data consist of the same pool of respondents, creating a three-year longitudinal file. This longitudinal file will be in the form of micro-data with reported data, administrative data, or synthetic data for variable items. The form will be decided upon in discussions with stakeholders.

As the table in Attachment A shows, the last data set deliverables from the 2004 SIPP Panel will be released in 2007 and early 2008 (February for the Wave 8 core file and December, January, and February for the topical modules). In addition, the Census Bureau will continue to produce analytical reports evaluating the 2004 SIPP Panel. Current SIPP P70 reports can be found on the Census website at http://www.sipp.census.gov/sipp/p70/p70s.html. This collection includes a series of report on the Dynamics of Economic Well-being (e.g., P70-105, P70-100).

The *Data Products Group* will work in parallel with the Prototype and Survey Content groups to determine the data products that will be produced from the new system. This determination will be based primarily on the needs of our data users, both internal and external and any disclosure issues identified by the Census Bureau’s Disclosure Review Board. While the Data Products Group has been established, work in earnest will not commence until the survey methodology has been finalized.

V. STAKEHOLDER SUPPORT

Stakeholder support is crucial to the success of this new system. We envision this as a collaborative effort between the Census Bureau and its many stakeholders during the lifetime of the project. Meetings with each individual agency to determine its needs are ongoing. As the development of the system progresses, we will continue to have these inter-agency meetings to discuss our progress as well as any issues that have arisen.

CONCLUSION

The structure of the interview, whether it requires personal visit interviewing, the number of re-interviews, whether movers are followed, and the requirement to measure program participation and eligibility must be balanced with the cost and sample size considerations. In addition, the timeliness of data production and release with respect to collection has long been an area where the SIPP program has faced criticism. Among the objectives for this re-engineering are improvements in the timeliness of data releases with respect to data collection, and improvement of processing efficiency by simplifying the structure of the files and the data collection. Several of the options being considered focus on only a single dynamics of economic well-being survey per ‘panel’ (that is, 2 years of data from a current survey and a third year from a new annual survey product.

2006 Current Survey	2007 Current Survey	2008 New Survey		
	2007 Current Survey	2008 Current Survey	2009 New Survey	
		2008 Current Survey	2009 Current Survey	2010 New Survey

However, there are still options being considered with multiple survey administrations per ‘panel’. These options, while providing more products and a mix of products, are very processing intensive and may once again interfere with the timely release of data. It is critical that this new system provide data in a timely manner at reduced cost, while continuing to provide the most important information to policymakers.

Attachment A

2004 SIPP Panel Data Products Release Dates

Wave	Core Release Date	Topical Module ™	TM Release Date
1	5/2006	Recipiency History	4/2006
		Employment History	5/2006
2	6/2006	Work Disability History	6/2006
		Marital History	6/2006
		Fertility History	6/2006
		Household Relationships	6/2006
		Migration History	6/2006
		Education and Training History	6/2006
3	7/2006	Medical Expenses/Utilization of Health Care	6/2006
		Work-related Expenses	6/2006
		Child Support Paid	6/2006
		Child Well-being	6/2006
		Assets, Liabilities, and Eligibility	9/2006
4	8/2006	Annual Income and Retirement Accounts	7/2006
		Taxes	7/2006
		Work Schedule	7/2006
		Child Care	8/2006
5	9/2006	Adult Well-being	3/2007
		School Enrollment and financing	4/2007
		Child Support Agreement	4/2007
		Support for Non-household Members	4/2007
		Functional Limitations and Disability ! Adult	4/2007
		Functional Limitations and Disability ! Child	4/2007
		Employer Provided Health Benefits	4/2007
6	10/2006	Assets, Liabilities, and Eligibility	5/2007
		Medical Expenses/Utilization of Health Care	5/2007
		Work-related Expenses	5/2007
		Child Support Paid	5/2007
7	11/2006	Annual Income and Retirement Accounts	8/2007
		Taxes	8/2007
		Retirement and Pension Plan Coverage	10/2007
		Informal Care-giving	10/2007
8	2/2007	Welfare Reform	12/2007
		Child Well-being	1/2008
		Child Care	2/2008