The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

LESSONS FOR TARGETED PROGRAM EVALUATION:

A PERSONAL AND PROFESSIONAL HISTORY OF THE SURVEY OF PROGRAM DYNAMICS

by

Daniel H. Weinberg *
U.S. Bureau of the Census

CES 07-24 **August, 2007**

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Bureau of the Census, 4600 Silver Hill Road, 2K132F, Washington, DC 20233, (301-763-1882) or INTERNET address sang.v.nguyen@census.gov.

Abstract

The Survey of Program Dynamics (SPD) was created by the 1996 welfare reform legislation to facilitate its evaluation. This paper describes the evolution of that survey, discusses its implementation, and draws lessons for future evaluation. Large-scale surveys can be an important part of a portfolio of evaluation methods, but sufficient time must be given to data collection agencies if a high-quality longitudinal survey is expected. Such a survey must have both internal (agency) and external (policy analyst) buy-in. Investments in data analysis by agency staff, downplayed in favor of larger sample sizes given a fixed budget, could have contributed to more external acceptance. More attention up-front to reducing the potentially deleterious effects of attrition in longitudinal surveys, such as through the use of monetary incentives, might have been worthwhile. Given the problems encountered by the Census Bureau in producing the SPD, I argue that ongoing multi-purpose longitudinal surveys like the Survey of Income and Program Participation are potentially more valuable than episodic special-purpose surveys.

* Chief, Center for Economic Studies and Chief Economist, U.S. Census Bureau. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion. The author wishes to thank Margo Anderson, Kathleen Creighton , Judy Eargle, Nancy Gordon, Donald Hernandez, Charles Nelson, Jennifer Rothgeb, and Stephanie Shipp for their helpful comments and suggestions, and for providing copies of historical documents. Contact information: <daniel.h.weinberg@census.gov>.

INTRODUCTION

The Survey of Program Dynamics (SPD) was an experiment in creating a national data set specifically for one purpose – evaluating the comprehensive welfare reform legislation enacted in August 1996. The use of those survey data by researchers has been disappointing – as noted by Haskins (2006, p. 418), "the [survey] had a problem with attrition and has unfortunately yielded less information than planned." This paper describes the evolution of the SPD and draws some lessons for the future about targeting program evaluation funds. It does not present evaluative findings about welfare reform from research that uses the SPD data.

First, it describes the historical origins of the SPD. It then focuses on the evolution of the SPD at the Census Bureau up to the time of enactment of the legislation. Then the paper covers implementation activities and difficulties. The final sections put the SPD in the broader context of welfare reform evaluation and discuss some of the lessons learned from the process of undertaking a new survey with little lead time and ambitious goals, and elicit some of the pitfalls to be avoided in undertaking new large-scale surveys.

I have undertaken an examination of SPD's history in the hope the reader (and I) can learn lessons from that story. However, more than a decade has elapsed since the first efforts to create this survey and errors due to faulty memory may have crept in. I hope my colleagues will revise and extend what is said here.[*]

ORIGINS OF THE SURVEY OF PROGRAM DYNAMICS

*. Though, naturally, any errors remain my responsibility.

This section discusses the three key progenitors of the SPD: (1) the early 1990s reexamination of the goals of the Survey of Income Program Participation (SIPP), (2) the Census Bureau's desire to provide the best possible data for public policy analysis based on experience with prior welfare reform evaluation efforts, and (3) the role of decision makers in facilitating the creation of relevant data.

Refocusing the SIPP

The original design of the SIPP as implemented in October 1983 was for overlapping panels, that is, a new longitudinal panel survey starting each year in February, consisting of eight waves, each with 4-month recall, each of 20,000 households, and thus covering 32 months for any one household member present at the first interview and followed for all waves. Before the first (1984) panel was even completed, however, budget cuts forced a reduction in its sample size, as well as for subsequent panels. By the end of the 1980s, it was clear that despite the Census Bureau's efforts, the SIPP budget would not be restored to its design sample size. The Census Bureau decided in 1989 that the 1989 panel should be canceled after three waves, and the funding thereby obtained be used to increase the size of the 1990 panel.

This radical step set off a reexamination of the goals of the SIPP, since funding could no longer be relied on to meet two prior but competing goals – to produce timely data for cross-section analysis (by combining data from overlapping panels for a given time period), and to produce the longer-term data needed for in-depth studies of longitudinal household composition, income, and transfer program dynamics. An internal team was formed to carry out an opinion survey (actually a census) of all known SIPP users, as well as a large number of potential SIPP

users.

Among the options presented by the Census Bureau's SIPP Alternative Design Group to the SIPP Management Redesign Team was "Option G" – a 52-month panel consisting of 13 interviews at 4-month intervals, followed by 6 annual interviews, covering a total of 10 years.[**,***] As noted in the report (U.S. Census Bureau 1992, p. 10), "designs with a panel length in the range of 8 to 15 years provide the time period necessary to answer questions about long-run patterns, such as causes, and consequences of poverty and dependence on government assistance programs."

The SIPP Management Redesign Team then formed the SIPP Extended Panel Work Group "to further explore the potential value and funding mechanisms for an annual follow-on survey that would effectively extend a SIPP panel to a ten-year horizon" (U.S. Census Bureau 1993, p. 1).[****] The work group concluded after extensive discussions with leading scholars that existing surveys do not "fill the need for long-term analysis of participation in welfare, health, and education/training programs, and the need for long-term analyses of the implications of participation in these programs for employment, income, and family and child well-being. ...The SIPP ... provides the best available point of departure for meeting the emerging federal need to measure the extent to which and the ways in which program participation produces desirable or

---

**. The author was a member of the SIPP Management Redesign Team.

***. Option D extended a 48-month SIPP panel to 12 years with annual interviews, and Option E included a supplemental 25-year panel of annual interviews. The idea for an extension of SIPP using annual interviews was brought to the Alternative Design Group (chaired by Enrique Lamas) by Donald Hernandez, a member of that group.

****. Members of the Extended Panel Work Group were Donald Hernandez (chair), Suzanne Bianchi, Enrique Lamas, and Louisa Miller, all members of the Alternative Design Group.

undesirable consequences, and the reasons for those consequences." (pp. 3-4).

Simultaneously, the Census Bureau commissioned the Committee on National Statistics, a part of the National Academy of Sciences, to study the future of SIPP. Their recommendations were published in Citro and Kalton (1993). While wide-ranging, the key recommendation was that "the length of each SIPP panel be increased from 32 to 48 months, ...that panels be introduced every 2 years rather than annually, ...retaining the 4-month recall length" (p. 4). However, they did not focus much attention on panels longer than 6 years, alluding only to one analyst's recommendation that a panel of 10 years be considered (Manski 1991).[*****]

Substantial debate ensued at the Census Bureau and the Office of Management and Budget's (OMB) interagency SIPP Advisory Committee. Eventually a new design (non-overlapping 4-year panels) was chosen and implemented in April 1996 after thorough questionnaire redesign involving cognitive pre-testing and field testing, conversion to computer-assisted personal interviewing, and selection of a new sample based on the 1990 Census (with an oversample of low-income households).[******] Difficulties for the SIPP did not end with implementation of the new design, however. A small 2000 panel was canceled after two waves (8 months) due to further funding difficulties and the desire for a larger 2001 panel, and unexpectedly high attrition led to a reduction in the panel length of the 2001 panel to 3 years

*****. The SIPP Extended Panel Work Group interviewed both the panel chair, Graham Kalton, and the panel study director, Constance Citro. Kalton and Citro both indicated that the panel felt the idea of a ten-year survey was outside their charge to investigate, but that the panel would not have opposed a plan to extend the SIPP with annual interviews. Both expressed their personal support for the idea, as did two other panel members, Robert Hauser and Patricia Ruggles. See U.S. Census Bureau (1993) pp. 7-8.

******. Implementation was delayed from the planned start date in February 1996 due to a lengthy government shutdown in late 1995 and early 1996. See Huggins and Fischer (1994) for further discussion of that SIPP redesign.

rather than 4. A new and larger panel was begun in 2004 and will continue until January 2008, albeit with a reduced sample size starting in October 2006 due to budget problems. An attempt by the Census Bureau to end SIPP data collection while the SIPP was redesigned yet again to both reduce costs and improve timeliness (the so-called Dynamics of Economic Well-being System), introduced as part of the President's fiscal year 2006 budget proposal, met with heated opposition by academics, think tanks, and eventually Congress and the plan was abandoned early in 2007. A new SIPP panel is scheduled to begin in February 2008.

Prior Welfare Reform Evaluations

The second progenitor of the SPD was my experience as an analyst (and later division director) at the Office of Income Security Policy, Office of the Assistant Secretary for Planning and Evaluation (ASPE), Department of Health and Human Services.[*******] ASPE convened a working group of analysts to discuss how to evaluate changes in the Aid to Families with Dependent Children (AFDC) program that were part of the Omnibus Budget Reconciliation Act of 1981.

Federal aid to families with children ("welfare") has a relatively long history. The AFDC program started as Title IV-A of the original Social Security Act of 1935 (then called Aid to Dependent Children).[********] Substantial changes to the AFDC program were made in 1967, 1981, 1984, 1988, and 1996.

*******. The Office of Income Security Policy is now the Office of Human Services Policy.

********. The post-1980 legislative history is from U.S. House of Representatives, Committee on Ways and Means 1996, pp. 517ff.

Concerns about growing welfare rolls led to the 1967 changes, which introduced a work incentive – an income disregard (when computing benefits) of the first $30 and one-third of any remaining earnings. The 1981 law ended that disregard after 4 months on a job, but the Deficit Reduction Act of 1984 reinstituted the $30 disregard for a year and allowed workers who lost eligibility to retain Medicaid for an additional 9 months (or up to 15 months at state option). The Family Support Act of 1988 introduced the Job Opportunities and Basic Skills (JOBS) program and permitted Medicaid and child care support to continue for families who lost AFDC eligibility because of earnings. JOBS required states to offer education, job skills training, job readiness, job development, and job placement, and also required two of the following: group and individual job search, on-the-job training, work supplementation, and community (or other) work experience.

It was a sobering experience for me to find the work group convened to evaluate the 1981 legislation stymied by lack of data. Research was eventually performed, but it consisted almost exclusively of either inference from repeated cross-sectional surveys not designed for welfare evaluation (mainly the Current Population Survey, see Moffitt 1986), or anecdotal inference from selected post-reform studies (see U.S. General Accounting Office 1985).

Yet the 1981 and 1984 changes were not enough for the states. They wanted more flexibility in program design and the Reagan Administration wanted to give it to them. Consequently, during the later years of the Reagan Administration, I was involved in designing evaluations for the welfare reform demonstrations approved by the Interagency Low Income Opportunity Advisory Board (see Fishman and Weinberg 1992). Obtaining good comparative data for the multiplicity of non-experimental demonstrations was key to good evaluation

planning. This was followed by my involvement in detailed planning for experimental evaluation of the JOBS component of the Family Support Act of 1988. Yet despite this new legislation, states continued to demand the flexibility to design their own welfare programs, particularly ones that emphasized work over welfare.

The possibility of a nationwide change in welfare (and federal health) programs re-surfaced as President Clinton took office in 1993, and the need for additional data to evaluate any changes arose as well. Experimental design was not possible, so national surveys to allow inference from longitudinal data and from state differences in implementation seemed valuable ways to proceed. Compared to 1981, however, there was now a survey environment – the SIPP – suitable for use as a pre-reform baseline.

But what sort of survey should be conducted? My summary of the November 1996 Conference on Evaluating Comprehensive State Welfare Reform described the consensus as the optimal design for evaluating welfare reform as "a pre-post study using a multiyear panel survey supplemented by examination of administrative data and some comprehensive site studies" (U.S. Census Bureau 1996c).********* Illustrative quotes include

- Robert Haveman: "...most of the answers to the questions of interest [about welfare reform] can be obtained through a well-designed, longitudinal sample surveys of households";
- Thomas Kaplan and Daniel Meyer: "annual data on the same families over a period of more than five years is desirable"; and

---

*********. The conference was sponsored by the University of Wisconsin Institute for Research on Poverty.

- Maria Cancian and Barbara Wolfe: "only a multi-year longitudinal design can assess the complex factors influencing individual life-course events."

I held this perspective as well in the early and mid-1990s.

Education of Congressional and Executive Staff

As one of the few management officials at the Census Bureau with evaluation experience (having moved there to be Chief of the Housing and Household Economic Statistics Division in 1989), I was in a unique position to inform both executive and legislative branch staff about the need for advance planning for evaluation and specifically about the need for good data.[**********] This educational effort received a good reception at the House Ways and Means Subcommittee on Human Resources, whose subcommittee staff worked together across party lines to provide data valuable for policy analysis, such as the "Green Book" produced under both Democratic and Republican leadership (e.g., U.S. House of Representatives, Committee on Ways and Means 1996, 1998).[**********] Census Bureau staff and I were able to discuss the importance of program evaluation with both Wendell Primus (majority staff director through 1993, later Deputy Assistant Secretary for Human Services Policy in ASPE from 1993-1996 and now a special assistant in House Speaker Pelosi's office),[**********] and Ron Haskins (minority staff director

---

**********. I was accompanied on some of these educational visits by Donald Hernandez, a member of both the SIPP Alternative Design Group and the Extended Panel Work Group.

**********. Also supportive was Paul Offner, of Senator Daniel Patrick Moynihan's staff. Senator Moynihan was a member of the Senate Finance Committee, with jurisdiction over welfare.

**********. After he resigned from ASPE in 1996 to protest the President's decision to sign the welfare reform legislation, Primus worked for the Center on Budget and Policy Priorities and then the Joint Economic Committee.

through 1993, then majority staff director 1994-2000, and now at the Brookings Institution).**********

Meanwhile, the Census Bureau developed a proposed set of topics in December 1992 and began discussions with analysts at ASPE and at the Department of Agriculture's Food and Nutrition Service (FNS, responsible for research on the food stamp program, was then Food and Consumer Services). After briefing both agencies, the Census Bureau began work on a Fiscal Year 1995 budget initiative to seek funds for a program to provide data to support evaluation of forthcoming reforms.

Both ASPE and FNS provided seed money to the Census Bureau to begin design of a questionnaire instrument to collect information expected to be of use in evaluating reform.********** The interests and needs of ASPE and FNS provided a valuable focus for design, as did a desire to track child well-being outcomes for several of those involved (particularly Haskins and Hernandez). This confluence of ideas between Congressional staff and executive branch bureaucrats was the third key progenitor of the SPD.

DESIGNING A POSSIBLE NEW SURVEY

After the report of the Task Force on National Health Care Reform received a chilly reception from the Congress and the public, it became clear that efforts in preparing data for

---

**********. One notable milestone was a February 2, 1995 letter from Daniel Weinberg to Ron Haskins giving historical precedents and funding estimates (U.S. Census Bureau 1995).

**********. The survey was named the "Survey of Program Dynamics" by April 14, 1993, primarily at my suggestion. The questionnaire design was led by the Census Bureau's Center for Survey Methods Research, particularly Jennifer Hess and Jennifer Rothgeb, and included consultations with both government and non-government policy analysts and survey methodologists.

policy evaluation ought to focus solely on welfare issues. A key memorandum of December 9, 1994 (U.S. Census Bureau 1994) laid out two goals of the SPD:

- "Providing information on spells of actual and potential program participation over a ten-year period;" and
- "Examining the causes of program participation and its long-term consequences on the well-being of recipients, their families, and their children;"

along with guidelines for content, preliminary data collection decisions, and a timeline. The plan was to design a questionnaire for a May 1996 pretest (using selected 1992 SIPP panel households as the sample), with the full survey to be administered in March through May 1997 to collect calendar year 1996 information from households who participated in the 1993 SIPP (who would then have been interviewed twelve times from February 1993 through January 1996).

Welfare reform legislation (H.R. 4) did not receive Congressional support in the Fall of 1995, and, as a consequence, Census Bureau, ASPE, and FNS staff understandably lost enthusiasm for completing a questionnaire that likely would never be used. As the final design period also coincided with the government shutdown of late 1995-early 1996, questionnaire design was stopped, and field testing was suspended "due to uncertainty about future funding" (U.S. Census Bureau 1996a). In effect, the questionnaire design effort was mothballed. All staff were reassigned to other projects.

IMPLEMENTATION OF THE SURVEY OF PROGRAM DYNAMICS

Surprisingly to many, welfare reform legislation was reintroduced in the early summer of

1996 in H.R. 3734 and passed by the Congress in August 1996.********** Key to the successful inclusion of funding for the SPD in the bill was a July 15, 1996 bipartisan meeting led by Ron Haskins of the House Ways and Means Subcommittee staff.********** The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (Public Law 104-193) not only replaced AFDC with the Temporary Assistance to Needy Families program, it also required (and funded) the Census Bureau to:

> "continue to collect data on the 1992 and 1993 panels of the Survey of Income and Program Participation as necessary to obtain such information as will enable interested persons to evaluate the impact [of the law] on a random national sample of recipients of assistance under state programs funded under this part and (as appropriate) other low-income families, and in doing so, shall pay particular attention to the issues of out-of-wedlock birth, welfare dependency, the beginning and end of welfare spells, and the causes of repeat welfare spells, and shall obtain information about the status of children participating in such panels."

The Census Bureau began planning this mandated survey at once. The basis for planning was a white paper prepared for the July 15, 1996 meeting (U.S. Census Bureau 1996b) that listed nine key design features (pp. 1-2):

- The survey should measure program eligibility, access, and participation for the full range of welfare programs.
- The survey should measure the money income, in-kind benefits, and services received from programs.
- The survey should measure employment, earned income, and income from other economic sources.

**********. For background information on welfare reform prior to 1996 and the legislative history of the final bill, see Wiseman (1996) and Haskins (2006).

**********. In addition to Ron Haskins and others from the Ways and Means Subcommittee majority and minority staff, also attending were Nancy Gordon and Daniel Weinberg from the Census Bureau; Michael Laracy and William O'Hare from the Annie E. Casey Foundation; Donald Oellerich, Wendell Primus, Matthew Stagner, and one other person from ASPE; and Howard Rolston from the Department of Health and Human Services Administration for Children and Families.

- The survey should measure family composition, and resources available from extended family members.

- The survey should be a large, longitudinal, nationally representative study that allows dynamic changes in each of these areas to be studied, and the interrelationships linking these changes to be identified.

- The survey should include baseline data for a period prior to the initiation of reforms, and continue to collect data throughout the period of reform to monitor the process of change.

- The survey should continue to collect data for the period after the reforms are implemented to assess short-term to medium-term consequences or outcomes.

- The survey should measure child outcomes in the areas of cognitive development, educational achievement, physical and mental health, and behavior problems, and for older children economic independence and training programs, work, income, fertility, and family formation. It is important to measure child outcomes because one major aim of the reforms is investment in children to help them become productive citizens as adults.

- The survey should measure key features of the environments of children, because it is through these intervening mechanisms that reforms will have positive or negative consequences for child outcomes. An assessment of how and why reforms produce such consequences will be necessary to allow future refinements to welfare programs.

Consider what was involved:

- Since the household universe was double that originally planned (two SIPP panels rather than one), decisions had to be made immediately about how to select a sample that could be interviewed within the budget being provided;

- Since households in the 1992 panel were last interviewed in October 1994 through January 1995, contact had to be reestablished quickly in order to avoid even more difficulty in finding sample members, suggesting the need to be in the field in the Spring of 1997;

- Since all work on a potential survey had stopped 9 months earlier, there were no people

available to plan or prepare, and hiring new staff experienced with longitudinal surveys is time consuming and difficult;

- Since questionnaire development had been aborted, a different survey instrument had to be chosen to collect 1995 data for the 1992 SIPP panel (if possible) and 1996 data for both panels;
- Since the change in welfare was highly visible and political, the Census Bureau had to deal with increased interest from many parties about the content of the survey, including about whether to instead use the funding to supplement other studies (e.g., the Panel Study of Income Dynamics); and
- The sudden appearance of $10 million per year in the last month of a fiscal year and for several future years meant that many bureaucratic issues involving the use of those funds had to be addressed.

In addition to these daunting initial conditions, there were major internal obstacles to success:

- Because of the Census Bureau's commitment to carrying out its responsibilities for data collection on numerous existing surveys vital to the national interest, such as the Current Population Survey (CPS) – the vehicle for collecting monthly employment and unemployment statistics – and because of its commitments to collecting data for other federal agencies sponsoring surveys, its Demographic Surveys Division was unable to designate any of its existing senior staff to run the new program.
- The Demographic Statistical Methods Division was uncomfortable with the 26.6 percent attrition rate from the initial selected sample inherited from the 1992-1993 SIPP panels.

Nevertheless, the decision was made by senior managers to limit the SPD sample to households completing both the first and last waves of the 1992 or 1993 SIPP in order to maximize the amount of data available for analysis and to reduce the number of households to be interviewed to a more manageable number, given the budget constraint.

- The Field Division was concerned about the logistics of finding the right materials for the expired SIPP panels in storage (for example, the paper control cards), the necessity of finding people not interviewed for quite some time (over 2 years for some), and the need to hire new and therefore inexperienced field representatives.

Facing these and other significant difficulties and a tight time frame, the Census Bureau's Demographic Programs Directorate decided to attempt a radical approach – establishment of an inter-directorate team, with representatives from all relevant divisions including the Field Division (implementing an initiative of the Principal Associate Director for Programs to break down internal "stovepipes").********** As I was by default the chief management "sponsor" of the survey within the Census Bureau, this team was placed in the Housing and Household Economic Statistics Division, instead of the Demographic Surveys Division (the usual home for survey operations).

The team was given many responsibilities, but to ease its burden somewhat, it was decided early on to administer the March 1997 CPS and its Annual Demographic Supplement to the SPD sample as a "bridge" between the old SIPP questionnaire and a new SPD

**********. The team was headed by Patricia Johnson.

questionnaire.********** Further, development of a "CPS look-alike" data file for users based on the 1992 and 1993 SIPP had already been contracted out to the Urban Institute by the Social Security Administration for its own use, so a file would eventually exist to which the 1997 SPD "Bridge" survey and later SPD data could easily be matched.

An extensive series of outside consultations was begun, focusing on the members of the SIPP Technical Working Group of the American Statistical Association's Survey Methods Research Section and members of the Census Bureau's Advisory Committee of Professional Associations. Among the issues discussed in these meetings were

- recontacting SIPP households and tracking movers;
- structuring possible monetary incentives for participation;
- criteria to establish for subsampling;
- dealing with non-respondent SIPP households and methods to reduce potential attrition bias (that is, concern that the remaining sample would not be representative);
- imputing for missing annual data if necessary (specifically 1995 data for 1992 SIPP panel households);
- including children who had not moved with SIPP sample households;
- achieving consistency of questions across surveys;**********
- weighting, including what year to use as a base year (due to concerns about births and

---

**********. The CPS supplement, now named the Annual Social and Economic Supplement, is now administered from February to April each year, and collects the information necessary to make annual estimates of income, health insurance coverage, and poverty.

**********. Kristin Moore, President of Child Trends Inc., a Census Bureau contractor working on the SPD, performed a valuable intermediary role among various survey sponsors to promote consistent child-oriented questions.

immigration) and the value of state-based weights;

- cognitive and field testing of a new questionnaire;
- using event history calendars;
- choosing between a content reinterview versus a reinterview solely for verification.

While the team was eventually successful in fielding a survey in the Spring of 1997, it would be an exaggeration to call the team fully successful. While the team leader and the other managers involved were experienced in many areas, they were inexperienced in survey operations and were unable to recruit sufficiently experienced staff. Expectations, roles, and boundaries for team members versus non-team staff were not made sufficiently clear. For an existing survey, roles and responsibilities have been clarified over many years. In contrast, for a new survey, especially one being conducted under a new organizational paradigm, issues that fell into unclear boundary areas were sometimes ignored for far too long – one "side" would unilaterally consider an issue "the other group's problem," while that other group might either implicitly assume it was being addressed, or even worse, because of inexperience, not even know it needed to be addressed.

Any new team is difficult to assemble quickly and needs time to figure out how to work together. Unfortunately, the one thing they did not have was time. Moreover, I (the team sponsor) also had little experience in the operational aspects of surveys, and could not easily recognize danger signals. When it became clear that the SPD team needed more experienced people in order to succeed, none were available due to competing priorities and a Census Bureau hiring freeze.

The 1997 SPD Bridge survey attempted to interview all sample persons in the 38,000

households that completed all waves of the 1992 and 1993 SIPP panels (73 percent of the original sample). The field staff interviewed 82 percent of those households (approximately 30,000) in May and June 1997. After the SPD Bridge survey was fielded, in June 1997 the team was disbanded and responsibility for SPD operations was transferred to the Demographic Surveys Division branch chief responsible for the SIPP.********** Expectations for the SPD were further clarified, monitoring was transferred to a inter-divisional steering committee, questionnaires and supplements continued to be designed and redesigned, and eventually the survey got on a relatively even keel (eventually the group responsible for the SPD became a separate branch).********** In effect, the new structure replicated traditional and successful survey operations at the Census Bureau. To the best of my knowledge, no inter-directorate team was ever later established to execute survey operations, though there continues to be widespread use of inter-divisional working groups for all surveys.

A new core SPD questionnaire based on the initial (1995) effort described above was developed for 1998 administration with the assistance of Child Trends Inc. One of the key decisions stabilizing the administration of the SPD was the decision to "freeze" the core content after 1998, and include only a limited number of topical modules. The SPD core instrument included retrospective questions for all people age 15 and older on jobs, income, and program participation, as well as detailed questions about children under 15. The 1998 and the 2001 surveys included a self-administered adolescent questionnaire; the 1999 and 2002 SPD included

**********. In Demographic Surveys Division, survey operations were directed first by Michael McMahon, and then by Patrick Benton.

**********. The steering committee was chaired by Charles Nelson and then Stephanie Shipp. Key members included Vicki Huggins and Robert Kominski.

"extended" measures of children's well-being; the 2000 SPD included a retrospective children's residential history.

But the intense pressure of late 1996-early 1997 and the inability of the Census Bureau to take on the responsibility of a major survey on a moment's notice had lasting effects. Tradeoffs were made early on in favor of maximizing the SPD's sample size instead of hiring of analysts to work with and analyze SPD data. Further tradeoffs were made emphasizing collecting new data over processing already collected data, since data not collected on time can never be collected while data not processed on time can be processed later. The survey never had sufficient funding or people to process the increasingly complex dataset, or to analyze the results. Finally, the staff working on the SPD, a one-time survey, left at a higher than normal rate to work on more permanent projects.

Questionnaire review and testing and data clearance fell on the shoulders of existing SIPP analysts, already burdened with clearance of 1996 SIPP panel data. Thus, not only was SIPP data review delayed even more than expected (recall that the 1996 panel was totally redesigned and transferred to computer-assisted interviewing which required a completely new processing system), SPD data review and clearance slowed to a crawl.

The initial focus was on gathering complete year-by-year data on all SPD households for later analysts to use in longitudinal models. Advisory group member Greg Duncan argued for a different conception of what analysts would use. Longitudinal year-by-year data were valuable, but analysts would also find it useful to compare outcomes for a particular year based on initial conditions for the same household. This conjecture called into question the Census Bureau's initial decision to interview only those who had completed the 1992 or 1993 panels.

As noted, the SPD had inherited a 26.6 percent sample loss rate from the 1992/1993 Survey of Income and Program Participation (SIPP) sample.********** In 1999, after 3 years of annual SPD interviewing, the SPD basic sample loss rate was up to 49.8 percent (see Table 1). From previous studies of SIPP sample loss, we know that households in and near poverty attrit at a higher rate than other households (see Lamas et al. 1994; Mack and Petroni 1994; Zabel 1994). Since poverty households were a key target population in the study of welfare reform, there was considerable concern about potential nonresponse bias.

After an experiment showing a positive effect of monetary incentives on reinterviewing attriters (providing the evidence required by OMB to concur), and the positive experience of two other longitudinal surveys in using incentives (the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth), through the hard work of Ron Haskins and the support of Representative Clay Shaw additional funds were appropriated to address this concern. An additional $1 million in FY 2000 allowed the Census Bureau to bring back a representative sample of households which had refused to be interviewed for the 1997 SPD and households which had moved and not been located earlier. This process reduced the sample loss rate to 44.3 percent. In FY 2001, the Census Bureau received an additional $5 million to bring back a representative sample of these types of households from the 1992 and 1993 SIPP; this reduced the overall cumulative sample loss rate after the 2001 SPD interview to a respectable 34.2

**********. The sample loss rate consists of cumulative noninterview rates adjusted for unobserved growth in non-interviewed units. The base for the calculation is the initial sample of occupied housing units judged in-scope (occupied by civilian non-institutionalized people). In contrast, some longitudinal surveys use the interviewed sample from their first administration as their base (that is, they start from 100 percent after one wave, while the 1992 and 1993 SIPP panels started at around 92 percent after one wave). (See U.S. Census Bureau 1999, for a comparison of the SIPP/SPD, Panel Study of Income Dynamics, and National Longitudinal Survey of Youth response rates based on consistent definitions.)

percent. Unfortunately, no further appropriation for the 2002 SPD interview year was made by Congress, and unexpectedly high data collection costs led to a premature end of the 2002 SPD interviewing, so the cumulative sample loss rate after the 2002 interview was back up to 47.0 percent.**********

SPD data processing activities ended in January 2004. The key data product of the program is a well-documented weighted longitudinal file linking annual core data for all 11 years of the SIPP-CPS-SPD marriage (1992-2002).********** This file is supplemented by minimally edited cross-section data from each survey year to allow users to match additional measures to the core data for more in-depth analysis of particular subjects. As of July 23, 2007, there are 253 Google Scholar citations to the "Survey of Program Dynamics", some of which are to Census Bureau documents describing rather than analyzing the survey. Future analysis of the 1996 reforms will be easier and more accurate than if such data had not existed.

LESSONS FOR WELFARE REFORM EVALUATION

Evaluations based on the SPD can be only a part of the portfolio of the evaluations of the effects of the 1996 welfare reform. Compared with the data available to evaluate the 1981

**********. During the 1999, 2000, 2001, and 2002 data collection periods, interviewers were allowed to use $40 incentives to maintain sample and to convert refusals. In 2000, they used a $100 incentive to convert refusal and mover households reintroduced from the 1997 SPD Bridge Survey; in 2001, they used a $100 incentive to convert refusal and mover households reintroduced from the SIPP 1992 and 1993 panels. Additionally, a $40 conditional incentive was used in 2001 for completion of the Self-administered Adolescent Questionnaire. (See U.S. Census Bureau 2004.)

**********. See the *Survey of Program Dynamics Users' Guide* (U.S. Census Bureau 2004) and *Technical Documentation* at the Census Bureau's SPD web site <http://www.sipp.census.gov/spd/>.

changes to welfare, though, policy analysts are in much better shape to evaluate the 1996 changes. The components of the evaluation portfolio for the 1996 welfare reform include extensive management-oriented case studies, a new cross-section survey (the National Survey of American Families, funded by several foundations), the 1996 panel of SIPP, continuation of pre-reform welfare waiver experiments in selected states, specialized evaluations, and welfare "leaver" studies; Besharov (2003) has a useful discussion of some of the key issues and approaches.********** *No one evaluation will provide "gold-standard" evidence on the effect of welfare reform*. Yet state variation and these many approaches may allow multiple perspectives that will allow policy analysts to eventually craft reliable estimates of those effects.

Congress decided that the underlying revolution in welfare embodied by the 1996 legislation should be continued, and in PL109-171, the Deficit Reduction Act of 2005, it reauthorized welfare reform for an additional five years.********** Data collection for the SPD ceased in 2002, and data dissemination ceased in 2004. "New" data collection related to welfare reform and child well-being has not continued. Instead, the Census Bureau worked with the Administration for Children and Families to determine that using the authorized funds to expand the 2004 panel of the SIPP was in the best interests of welfare program evaluation. The funds were targeted on increasing the SIPP sample in states where sufficient additional sample could be added to improve the reliability of estimates for the low-income single parent households and

**********. See in particular the chapter by Peter Rossi, "Assessing Welfare Reform's Impact", and the comments on Rossi's 1999 conference paper that was the basis for that chapter, by Daniel Weinberg and Stephanie Shipp on the SPD, and by Kenneth Finegold and Fritz Scheuren on the National Survey of American Families, both published in the conference volume (Besharov 2002) but not in the 2003 book.

**********. PL109-171 added additional child care funding to support welfare reform, and funds for "healthy marriage" and "responsible fatherhood" programs.

their children in those states. In addition, the existing topical module on child well-being was moved to the third and ninth waves of the SIPP in order to ensure that data were available in a timely manner to assess changes in child well-being.

LESSONS FOR NEW LARGE-SCALE SURVEYS

Despite the problems identified above, all the data were delivered as envisioned, albeit not as quickly as might be desirable nor were they as thoroughly reviewed and edited as for the typical ongoing Census Bureau survey. I believe there are lessons to be learned for future implementation of new large-scale surveys.

First, the data collection agency must make it very clear to Congress what can and cannot be accomplished for the amount of funding provided. Time is necessary to assemble an adequate and experienced staff, to develop and test an appropriate questionnaire, and to at least begin if not complete the development of a processing system for the data as they come out of the field. In the case of the SPD, if the original timeline and sampling strategy had held, many of the organizational problems the Census Bureau faced would likely have been largely avoided. For example, if welfare reform had passed in late 1995, a proper questionnaire could have been completed for 1997 and last-minute administration of the March CPS would have been avoided. Eighteen months seems to me a *minimum* amount of time to field a new survey based on an existing survey, though more time ought to be taken for thorough cognitive and field pretesting or for newly developed survey instruments. For example, planning for the SIPP began in 1975, with field testing begun in 1978 and the first large-scale survey planned for 1981 (budget issues delayed its start until October 1983). Planning and development for the American Community

Survey began in earnest in 1994, with large-scale implementation more than a decade later – in January 2005 (with full implementation, including people in group quarters, in January 2006).

Second, more thought ought to have been given to the allocation of funding between field operations and headquarters production. The Census Bureau gave the Congress an estimate of $9 million per year to continue to interview all the households in the 1993 SIPP. Instead, the Congress appropriated $10 million to continue to interview both the 1992 and 1993 panels, and it gave the first tranche of this money in the last month of fiscal year 1996. To maximize the sample size, the Census Bureau limited the number of professional staff assigned to work on the survey, and in consequence had to rely on the policy analysis community to handle data analysis. Even so, the sample had to be reduced after the 1997 Bridge survey when selective subsampling was used to retain all the households expected to be most affected by welfare reform. It might have been a wiser decision to accept a smaller sample and fewer field interviewers in order to have a more robust survey and analytical staff. Learning a new survey is time-consuming for an academic or policy analyst, and there was little motivation on their part to invest in that education.

Third, internal buy-in is as important as external buy-in. Additional experienced operations and analytical staff should have been assigned to the SPD at its beginning, but the survey's quick start and the Census Bureau's desire not to disadvantage long-standing existing surveys, such as the CPS and the SIPP, by reassigning existing staff was perhaps more the agency's long term interest. Experienced operations staff would have improved both quality and timeliness of the SPD. A professional analytic staff can focus data collection, data processing, and data review on the key aspects of a new survey. They can also market the SPD at key

professional conferences to popularize the dataset by presenting analyses of the data and insights on how to use the data.********** The Census Bureau underestimated the difficulties that policy analysts would face in analyzing a new data set, albeit one with a familiar "CPS look-alike" structure. (Indeed, the Census Bureau underestimated the difficulties that an externally determined file structure would have for its own data preparation activities.) Yet, in my opinion, the evaluation community has underestimated the potential of the SPD to answer important policy questions, and under-invested in learning about the dataset.

Fourth, the extent and deleterious effects of sample attrition were underappreciated. Part of this problem resulted from an unexpected secular increase in nonresponse to government surveys across the board. But following nine or ten waves of the SIPP with up to six additional waves of the SPD was unexpectedly harsh on the respondents. Special appropriations had to be made to add to SPD resources and special permission had to be obtained from OMB to allow the Census Bureau to use monetary incentives, and to reinterview prior nonrespondents dropped along the way without jeopardizing sample size. Longitudinal surveys are hard to do and adequate funding is necessary.

Yet despite the missteps, I am still firm in my belief that longitudinal surveys are a valuable part of the portfolio of evaluation tools needed to assess the impacts of large, national changes, particularly ones where experimental evaluations are not possible. In a review of Moffitt and ver Ploeg (2001), I said that "I endorse the multiplicity of valid approaches to

**********. Among the attempts at outreach were Kominski and Weinberg 1997, Weinberg 1997, Weinberg et al. 1997, Weinberg et al. 1999, Hess 2001, Fronczek et al. 2002, Hofferth 2002 and a special brochure developed for distribution at professional meetings (U.S. Census Bureau 2002). The Census Bureau also sponsored a "small grants" competition to use SPD for analysis and published some internal data analyses (e.g., Loveless and Hisnanick 2003; Overturf and Downs 2003).

evaluation. No one approach is likely to be the final word." (Weinberg 2002, p. 713) and I still agree with that statement. On the other hand, I now would modify the prescriptive part of my 2002 statement that "Special purpose data collection can *and perhaps should* be part of every major reform legislation." (p. 714, italics added). What has become clear to me in the past decade is that there is more value in having an ongoing large multi-purpose longitudinal household survey like the SIPP to supplement existing cross-section surveys than in developing a special-purpose survey to evaluate a particular program change.**********

**********. I also recommend that any new multi-purpose survey if possible have an overlapping design so that major program changes do not fall "on a seam" between abutting panels.

REFERENCES

Besharov, Douglas J. (ed.) *Family Well-Being After Welfare Reform.* Maryland School of Public Affairs, Welfare Reform Academy, 2002.

Besharov, Douglas J. (ed.) *Family and Child Well-Being After Welfare Reform.* New Brunswick, NJ: Transaction Publishers, 2003.

Citro, Constance F. and Graham Kalton (eds.). *The Future of the Survey of Income and Program Participation*. Washington, D.C.: National Academy Press, 1993.

Fishman, Michael C. and Daniel H. Weinberg. "The Role of Evaluation in State Welfare Reform 'Waiver' Demonstrations" in C. Manski and I. Garfinkel (eds.) *Evaluating Welfare and Training Programs*, 1992.

Fronczek, Peter, Patricia Doyle, Robert Kominski, and Charles Nelson. "Complex Problems – Complex Data: Welfare Reform and the Survey of Program Dynamics." Paper prepared for the American Statistical Association meetings, August 2002.

Haskins, Ron. *Work over Welfare: The Inside Story of the 1996 Welfare Reform Law.* Washington, DC: Brookings Institution Press, 2006.

Hess, Jennifer. *Preparing to Measure Welfare Reform Using the Longitudinal Survey of Program Dynamics: 2001.* SPD Analytic Reports SPD-2001-1. U.S. Census Bureau, August 2001.

Hofferth, Sandra L. "Use of the Survey of Program Dynamics for Studying the Implications of Welfare Reform for Families and Children." Child Trends Inc. April 16, 2002; also published as U.S. Census Bureau Report SPD-2002-1.

Huggins, Vicki J. and Donald P. Fischer. "The Redesign of the Survey of Income and Program Participation." SIPP Working Paper No. 203. Paper prepared for the American Statistical Association meetings, August 1994.

Kominski, Robert A. and Daniel H. Weinberg. "Tracking Devolution: The Survey of Program Dynamics." Paper prepared for the Population Association of America meetings, May 1997.

Lamas, Enrique, Jan Tin, and Judith Eargle. "The Effect of Attrition on Income and Poverty Estimates from the Survey of Income and Program Participation." Paper presented at the Conference on Attrition in Longitudinal Surveys, Washington, DC, February 1994.

Loveless, Tracy A. and John J. Hisnanick. *Receipt of Rental Subsidies: 1993 and 1999.* U.S. Census Bureau SPD Analytic Reports SPD-2003-1. U.S. Census Bureau, October 2003.

Mack, Stephen and Rita Petroni. “Overview of SIPP Nonresponse Research Data.” SIPP Working Paper No. 201 (formerly No. 9414). September 1994.

Manski, Charles F. “Prepared Statement” in *Welfare Dependency*. Hearings before the Subcommittee on Social Security and Family Policy of the U.S. Senate Committee on Finance. Washington DC: Government Printing Office, 1991, pp. 127-129.

Moffitt, Robert A. “Work Incentives in the AFDC Program: An Analysis of the 1981 Reforms.” *American Economic Review*, May 1986.

Moffitt, Robert A. and Michele Ver Ploeg (eds.). *Evaluating Welfare Reform in an Era of Transition*. Washington, DC: National Academy Press, 2001.

Overturf, Julia V. and Barbara Downs. “Adolescent Behavior and Family Relationships.” Paper prepared for the Population Association of America meetings, May 2003.

U.S. Census Bureau. Memorandum: “Report on SIPP Design Options” to the Survey of Income and Program Participation Senior Management Redesign Team from the Alternative Design Group. July 27, 1992.

U.S. Census Bureau. Memorandum: “Final Report of the Group” to the Survey of Income and Program Participation Senior Management Redesign Team from the SIPP Extended Panel Work Group. March 31, 1993.

U.S. Census Bureau. Memorandum: “Goals of the Survey of Program Dynamics” to Survey of Program Dynamics Work Group from Daniel H. Weinberg, for the Survey of Program Dynamics Steering Committee. December 9, 1994.

U.S. Census Bureau. Letter to Ron Haskins, House Ways and Means Committee, from Daniel H. Weinberg, Chief, Housing and Household Economic Statistics Division. February 2, 1995.

U.S. Census Bureau. Memorandum: “Survey of Program Dynamics (SPD)” to Nancy M. Gordon, Associate Director for Demographic Programs from Richard R. Storm, SPD Team Leader. March 19, 1996[a].

U.S. Census Bureau. “A Survey of Program Dynamics for Assessing Welfare Reform.” July 15, 1996[b].

U.S. Census Bureau. Memorandum: “Institute for Research on Poverty Conference on Evaluating Comprehensive State Welfare Reforms” to Survey of Program Dynamics Team and Steering Committee from Daniel H. Weinberg, Chief, Housing and Household Economic Statistics Division. November 27, 1996[c].

U.S. Census Bureau. Memorandum: "Comparing Response Rates for SPD, PSID, and NLSY" to Daniel H. Weinberg, Chief, Housing and Household Economic Statistics Division, from Donald J. Hernandez, Special Assistant. January 28, 1999.

U.S. Census Bureau. *Survey of Program Dynamics: Understanding Welfare Reform.* Report HHES/02-SPDWFB. Washington DC, August 2002.

U.S. Census Bureau. *Survey of Program Dynamics Users' Guide (Revised).* Report DSD/04-2879. Washington DC, March 2004.

U.S. General Accounting Office. *An Evaluation of the 1981 AFDC Changes: Final Report.* Report PEMD-85-4. July 1985.

U.S. House of Representatives, Committee on Ways and Means. *1996 Green Book: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means.* Committee Print WMCP 104-14. November 1996.

U.S. House of Representatives, Committee on Ways and Means. *1998 Green Book: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means.* Committee Print WMCP 105-7. May 1998.

Weinberg, Daniel H. "A Survey of Program Dynamics for Assessing Welfare Reform," *[Institute for Research on Poverty] Focus*, Spring 1997.

Weinberg, Daniel H. "Book Review: *Evaluating Welfare Reform in an Era of Transition*." *Journal of Policy Analysis and Management* v. 21 no. 4, Fall 2002, pp. 712-715.

Weinberg, Daniel H., Vicki J. Huggins, Robert A. Kominski, and Charles T. Nelson. "A Survey of Program Dynamics for Evaluating Welfare Reform." *Proceedings of the Statistics Canada Symposium 97* "New Directions in Surveys and Censuses." November 1997.

Weinberg, Daniel H., Patricia Doyle, Arthur Jones Jr., and Stephanie Shipp. "Measuring the Impact of Welfare Reform with the Survey of Program Dynamics." *Proceedings of the American Statistical Association*, August 1999.

Wiseman, Michael. "Welfare Reform in the United States: A Background Paper." *Housing Policy Debate* vol. 7, no.4, 1996.

Zabel, Jeffrey. "An Analysis of Attrition in the PSID and SIPP with an Application to a Model of Labor Market Behavior." SIPP Working Paper No. 191 (formerly No. 9403). U.S. Census Bureau, 1994.

Table 1. Sample Loss Rates for the 1992 and 1993 Survey of Income and Program Participation (SIPP) Panels and the Survey of Program Dynamics (SPD)

Interview	Eligible households	Households in Sample	Interviewed households	Sample Loss Rate (non-interviews as percentage of sample)
1992/1993 SIPP	48,109	38,345[a]	35,291	26.6
1997 SPD "Bridge"	51,320	34,609[b]	30,125	41.3
1998 SPD Basic Sample	32,800[c]	19,129	16,395	50.0
1999 SPD Basic Sample	33,200	19,303	16,659	49.8
2000 SPD Basic Sample	33,600	19,802	16,845	49.9
2000 SPD Basic Sample + 1997NI	33,600	23,258	18,716	44.3
2001 SPD Basic Sample	34,000	20,185	16,964	50.1
2001 SPD Basic Sample + 1997 NI	34,000	23,801	19,236	43.4
2001 SPD Basic Sample + 1997 NI + 1992/1993 SIPP NI	34,000	29,341	22,340	34.2
2002 SPD Basic Sample	34,400	NA	10,765	54.4[d]
2002 SPD Basic Sample + 1997 NI	34,400	NA	11,693	50.5[d]
2002 SPD Basic Sample + 1997 NI + 1992/1993 SIPP NI	34,400	23,600	12,496	47.0[d]

Notes: NI = noninterview subsample. NA = not available.
a. As of the last wave (tenth interview, except ninth interview for one-quarter of the 1992 panel).
b. Only those households interviewed in the (first and) last wave of the 1992 or 1993 panels were eligible for the SPD Bridge survey that used the March 1997 Current Population Survey questionnaire.
c. Households interviewed in the 1997 SPD and selected during subsampling were eligible for the 1998 SPD. The numbers of eligible households (e.g., 32,800 for the 1999 SPD) are estimates based on the assumption that sample loss is uniformly distributed among the subsampled groups. The subsampling strategy for the 1998 SPD was 100% of households with incomes below 150 percent of poverty, 100% of households with children and incomes 150-200 percent of poverty, 90% of households with children and incomes above 200 percent of poverty, 80% of households without children and incomes 150-200 percent of poverty, and 27% of remaining households and of sample cases lost due to institutionalization (see U.S. Census Bureau 2004, for more details).
d. Because no extra funds were appropriated for the 2002 NI samples, all samples were reduced by one-third and the sample loss rates are calculated on a base of 23,600 eligible households.

Sources: U.S. Census Bureau 2004; Email message S.S. Ung to P.J. Benton "Re: SPD Sample Loss..." October 14, 2004 [for corrections].