

**Description of plans for a SIPP calendar validation study:
Study design and analysis**

Jason M. Fields, Housing and Household Economic Statistics Division, US Census Bureau
Jeffrey C. Moore, Statistical Research Division, US Census Bureau

This report is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Abstract

Plans for the Census Bureau's re-engineered Survey of Income and Program Participation (SIPP) program include use of event history calendar (EHC) interviewing methods, and (assuming a favorable research outcome) a 12-month, calendar-year reference period, in place of a standard questionnaire approach with a sliding 4-month reference period. This paper describes the first field test research project to compare the quality of the data obtained under the two approaches. The essential feature of the research is a small-scale field test, in early 2008, of a prototype paper EHC questionnaire, covering calendar year 2007, administered to expired 2004 panel SIPP households who will have already reported about calendar year 2007 via their final three waves of SIPP interviews. Analysis will focus on a comparison between the two interviewing methods of the reporting of key characteristics (e.g., participation in programs, jobs/businesses, and health insurance coverage), their start and stop dates, and (where relevant) income amounts. Because little is known about how EHC methods are actually put into practice in the field, the 2008 study will also employ a variety of additional evaluations -- interviewer and respondent debriefings, observations, analysis of recorded interviews, etc. -- directed toward a better understanding of the EHC interview process. Subject to available funding, the field test will be administered in one or two states, most likely IL and TX. Administrative records data to validate program participation from the two survey based estimates are in the process of being obtained. Following the survey based analyses; validation evaluations will be conducted with these records.

Overview

The US Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP) to accomplish several goals, including reducing burden on respondents, reducing program costs, improving accuracy, improving timeliness and accessibility, and improving relevance. The main objective of the SIPP has been to provide accurate and comprehensive information about the income and program participation of individuals and households in the United States. The survey's mission is to provide a nationally representative sample for evaluating: 1) annual and sub-annual income dynamics, 2) movements into and out of government transfer programs, 3) family and social context of individuals and households, and 4) interactions among these items. The survey re-engineering of SIPP pursues these objectives in the context of several goals - cost reduction and improved accuracy, relevance, timeliness, and accessibility. The SIPP collects detailed information on cash and non-cash income (including participation in government transfer programs) three times a year, and detailed data on taxes, assets, and liabilities are collected annually. A major use of the SIPP has been to evaluate the use of and eligibility for government programs and to analyze the impacts of options for modifying them.

A key component of the re-engineering process involves the proposed shift from the every-four-month data collection schedule of traditional SIPP to annual data collection in the re-engineered survey. To accomplish this shift with minimal harm to data quality, the Census Bureau proposes to employ event history calendar (EHC) methods to gather SIPP data (Fields and Callegaro, 2007). Belli (1998) provides a strong theoretical rationale for the use of EHC methods, and their likely superiority to more traditional survey instruments using a standard question-by-question approach. Most existing EHC evaluations are consistent with the hypothesis of improved data quality – by improvements in the ability of respondents to integrate memory across topic areas, and retrieve related information in a more natural autobiographical manner. The research base is somewhat limited in terms of strong quantitative evaluations of theory-based predictions. Most studies have focused on the use of comparable survey recall periods and evaluated strictly the survey method. Thus, concern lingers about the data quality implications for the topics covered in SIPP of the shift from a four-month recall period to a one-year recall period.

Background

The event history calendar (EHC) is a survey methodology that has been successfully employed since the 1960's to assist interviewers in collecting detailed data with long recall periods (Belli, 1998; Belli, Shay, and Stafford, 2001; Callegaro 2007). Although never implemented as a production instrument the Census Bureau and SIPP researchers have experience with EHC instruments. In the late 1980's an EHC was field tested with SIPP in the Chicago region (Kominski, 1990). In the end this test was not implemented as a production component because there were too many concomitant changes required to integrate it into the program. In the late 1990's, EHC instruments began to be developed as electronic instruments, significantly easing some of the issues associated with retrieving and coding the data collected with this tool.

The EHC methodology helps interviewers and respondents by allowing recall of information in a more natural “autobiographical” manner. Each spell can happen before, after, or at the same time as another spell. For example, a residence change can and in many cases occurs contemporaneously with a change in employment. The entire process of compiling the calendar focuses, by its nature, on coherence, consistency, sequential order, and attempts to correct for missing data (Belli, 1998). By coherence, it is intended that some events are less likely to occur together, such as having three jobs at the same time. Although that might be the case, it is also possible that the respondent made a mistake in the location of the jobs in time. The calendar instrument visualizes better than a traditional question-list instrument events in the time line suggesting possible inconsistencies. The sequential nature of the EHC is revealed by the fact that an event should happen after the preceding one and before the following one. For example if the respondents are unemployed, they will then look for a job, and when found, become employed. The time line also highlights to the interviewer missing data in a more prominent way than a traditional question-list. This is because each event in a time line should be adjacent to another event without any time unaccounted for (something must have happened in each time frame). In case of time unaccounted for, the interviewer can probe to investigate what happened in that time frame.

Event history calendar instruments have been evaluated on numerous occasions most involve comparison with a previously collected questionnaire administered to the same respondents (Freedman, et al. 1988; Caspi, et al. 1996; Ensel, et al. 1996; Belli, Shay, and Stafford, 2001). While each of these studies utilize reinterview comparisons, and Belli and colleagues (2001) explicitly designed a test-retest experiment where prior respondents were reinterviewed by two treatment groups, one with a two year EHC and one with a traditional instrument. The findings from all of these studies suggest that the EHC methodology yields high levels of agreement with the earlier data for most domains. Belli and colleagues (2001) find better recall for the same length recall period using the EHC for most topics and where the EHC was not an improvement it did not differ from the question-based interview. While AFDC and Food Stamps were among the categorical topics that did not differ by method, these subjects were not the primary focus of the survey or of training. The null finding in the PSID study for whether these programs could be collected better by EHC is one of the reasons the field test research described here is a necessary decision point for using an EHC based instrument in SIPP. Yoshihama and colleagues (2005) studied intimate partner violence reporting, and compared the reporting in two samples of women, one interviewed with traditional interviewing and one with a life history calendar. In this study there was no reinterview comparison to previous data but an assumption that, given the sensitive nature of the events and the lifetime perspective, more reported events indicated a more effective interviewing technique. In this case, the life history calendar provided significantly higher levels of reporting about intimate partner violence compared with traditional interviewing.

While agreement with a prior data collection, or more reporting of certain events can be a good indication that the EHC methodology is an improvement over question list collection of events, it

is not always a clear indication. More reported events are often treated with skepticism especially in the case of multi-wave surveys. Since the very beginning, researchers have considered it almost axiomatic that the amount of change measured between interview waves is overstated. Collins (1975), for example, speculates that between two-thirds and three-quarters of the observed change in various employment statistics (as measured in a monthly labor force survey) were spurious; Polivka and Rothgeb (1993) estimate a similar level of bias. Michaud et al. (1995) describe apparent change in income across successive survey waves as “grossly inflated” [p13]; similarly, Lynn and Sala (2006) label the amount of change they observe from one survey wave to the next in various employment characteristics as “implausibly high” [p8]; see also Cantor and Levin (1991), Hill (1994), Hoogendoorn (2004), and Stanley and Safer (1997). Recent research also shows that EHC collected data can decrease seam effects between waves compared with traditional questionnaire interviewing, and also potentially reduce seam effects between different components of an the same instrument (Callegaro, 2007). Additional work is needed to evaluate whether additional reductions in seam biases can be realized by combining dependent interviewing with EHC methodologies.

Other researchers have focused on the other side of the equation – the understatement of change within an interview wave – sometimes called “constant wave responding” (Martini, 1989; Rips, Conrad, and Fricker, 2003; Young, 1989). Moore and Marquis (1989), using record check methods, confirm that both factors – too little change within the reference period of a single interview, and too much at the seam – operate in concert to produce the seam effect. Kalton and Miller (1991) offer supporting evidence for that assessment, as does LeMaître (1992).

To help disentangle differences and reach substantiated conclusions about which methodology captures events appropriately, we have included an additional validation component to determine the accuracy of both the EHC responses and of the comparison SIPP data. None of the prior EHC evaluation reinterview studies described above were able to include a validation component, and none focused explicitly on the types of program related transitions of particular concern to the SIPP stakeholders. As part of our decision to move toward the collection of SIPP via an event history calendar, discussed in Fields and Callegaro (2007), it was clear that we would need to evaluate the EHC methodology for use on traditional SIPP concepts. The evaluation would place, perhaps, the strongest demands possible on the methodology – could an EHC with a one-year recall period provide data of comparable quality to that from production SIPP interviews with repeated question based interviews with 4-month recall periods.

Research Plan

This research paper describes the plans for the first SIPP reengineering evaluation and validation field test. The essential feature of the research is a small-scale field test, in early 2008 of a prototype EHC questionnaire, covering calendar year 2007, administered to expired 2004 panel SIPP households who will already have reported about calendar year 2007 via their final three waves of SIPP interviews (see Figure 1).

Figure 1.

SIPP 2004 PANEL REFERENCE PERIOD MONTHS IN CALENDAR YEAR 2007 BY ROTATION GROUP									
CALENDAR MONTH		ROTATION GROUP							
		1		2		3		4	
		Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month
2006	October	W10	W9						
	November			W9					
	December			W10		W9			
2007	JANUARY	W11			W10		W10	W9	
	FEBRUARY		W10						
	MARCH			W10					
	APRIL	W12		W11		W10	W11		
	MAY				W11			W10	
	JUNE		W11						
	JULY	W12		W12			W12		
	AUGUST					W11			
	SEPTEMBER			W12				W11	
	OCTOBER	W12	W12		W12		W12		
	NOVEMBER				W12				
	DECEMBER					W12			
2008	January							W12	
** FEBRUARY 2008 – START OF NEW 2008 PANEL **									

Our analysis will focus on a comparison between the two interviewing methods focusing on the reporting of key characteristics (e.g., participation in programs, jobs and businesses, health insurance coverage, school enrollment, and residences), their start and stop dates, and (where relevant) income amounts. Because little is known about how EHC methods can be put into practice in the field on a large scale federal survey, the 2008 study will also employ a variety of additional evaluations – interviewer and respondent debriefings, observations, analyses of recorded interviews, and training evaluations – directed toward a better understanding of the EHC interview process. The qualitative information gained from these observations will help to refine the training and identify problems that will need to be addressed before the 2009 dress rehearsal is fielded. We also hope that these qualitative methods can be useful in understanding any differences in the quantitative data collected by the EHC from the comparison data. These

evaluations will help to differentiate procedural issues that can be corrected through training from inadequacies in the instrument or methodology.

Subject to available funding, the field test will be limited to one or two states, Illinois (IL), Maryland (MD) and/or Texas (TX) are the possible states considered for this test. These states were chosen for ease of administration and, primarily to facilitate the use of administrative record data for a more rigorous data quality validation assessment for selected characteristics. These states are ideal test areas for this evaluation, with diverse populations and interviewing situations. There are sufficient cases from SIPP 2004 in these areas and there is solid groundwork in place to establish the necessary agreements to utilize administrative records in the validation step to the analysis. Table 1 presents the current households available to be interviewed in each area (and Maryland), and identifies them as continuing (Wave 10) households or sample-cut (Wave 8) households. If on average there are two adult respondents per household, interviewing about 1,000 households will generate nearly 2,000 individual EHC records for analysis. About half the available households are in the Wave 10 (reinterview sample), which would be the base for the evaluation comparing the EHC responses to their SIPP 2004 responses. The cases in Texas are a metropolitan subset of all the 2004 SIPP cases in Texas. We chose to focus on respondents in metropolitan areas to maximize the program cases available for evaluation and create a more cost efficient sample in Texas.

Table 1. 2008 Field Test -- Approximate Number of Available Cases

SIPP 2004 Available Cases (1)	Illinois	Texas	Maryland
	Available Households	Available Households	Available Households
Total households	936	1048	884
Wave 10 completed households	508	614	268
Wave 8 reduced households	428	434	616

Source: Survey of Income and Program Participation - 2008 Re-engineering field test 1.

Notes: (1) Households were selected for interview in the field test from those completing interviews through Wave 10 in Illinois and in four metropolitan areas of Texas.

(2) If a selected address interviewed for the Event History Calendar test does not include any SIPP 2004 respondents we will utilize the cases as a type of 'un-primed' replacement households.

Additionally, this test will provide these regional offices, their management, and field representatives with experience with the EHC survey methodology. This experience will be invaluable as we transfer what we learn about training interviewers on this first field test EHC to more regional offices for the full 2009 dress rehearsal.

There are several fundamental assumptions that need to be discussed as we move forward with the plans for the analysis of the reinterview and validation field test:

1. We assume that the results from a paper questionnaire with minimal content can be generalized to an automated EHC questionnaire with full content. There are limited ways that we can control the experiment to ensure that this assumption is upheld.
2. We assume that the results from continuing SIPP respondents are generalizable to respondents who would be new respondents in a new panel with no prior SIPP experience. As with any reinterview study, this assumption is key. Additionally, because 2007 is at the end of a long SIPP panel the comparison group also has a significant amount of experience with the SIPP instrument and SIPP topics, these are about as well trained as any survey respondents could be. Unlike the Belli et al. (2001) study, there was never any intent to replace the four-month SIPP with a twelve-month question list instrument, and as such, the split sample reinterview experiment design was not directly transferable.
3. We assume that the lessons learned and materials developed for training interviewers to administer the EHC 2008 field test can be adapted and improved for the 2009 dress rehearsal and then again for production SIPP. Further we assume that we will be successful in training the Census Bureau's field staff to administer an EHC instrument and develop the necessary probing and cueing techniques required to record high quality EHC data.
4. We assume that the results from a survey in limited areas and among a non-representative sample are generalizable to a national sample. We assume that the biases incurred due to non-response (through differential attrition and sample aging) in the source SIPP households can be described and as these issues do not preclude use of the later waves of SIPP, they do not invalidate the study findings based on these available SIPP households.
5. We do not assume that a difference between a SIPP and EHC estimate, or differential reporting of a status/transition indicates which method is more accurate. The validation activities will help to reconcile those questions. The first evaluation component is to describe the differences and levels of agreement on a number of dimensions. Issues related to seam bias and changing respondents are potential causes for erroneously high transitions in SIPP that could lead to false conclusions in a more is better framework for transitions.

One aspect of the study design is the fact that the SIPP respondents' EHC reports will be "primed" by having just completed three waves of SIPP interviews covering the same time period. Certainly, the experience of having been SIPP respondents will predispose these respondents to being able to accurately recall they type of information we have included in this test, just as their experience answering SIPP for the past three years improves their ability to navigate the complicated concepts in the SIPP instrument. This "priming" would be a significant problem if there were no plan to evaluate its effect. However, this study will yield data about the

The first comparison that we will be making is simply to assess the recording of events in EHC vis-à-vis the SIPP control data (SIPP vs. EHC (primed)). Responses to the 12-month EHC will be compared with the same respondents' SIPP interview reports covering the same calendar year. Missed events in one or the other interview method are likely evidence of reduced data quality. The events being evaluated include (Key SIPP Variables Involved – Public Use Names):

1. Residential Moves (SHHADID, TFIPSST, TMETRO, RHCHANGE, EPUBHSE, EGVTRNT, EWRSECT8)
2. School Enrollment (RENRLMA, EERLM, EENLEVEL, EEDUCATE)
3. Labor Force (EBNO1, EBNO2, TBSOCC1, TBSOCC2, EENO1, EENO2, TJBOCC1, TJBOCC2, RPYPER1, RPYPER2, TPMSUM1, TPMSUM2, RMERS, ELAYOFF, ELKWRK, RWKESR1, RWKESR2, RWKESR3, RWKESR4, RWKESR5, TFUNEMP)
4. Workers Insurance Programs (ER05, ER06, ER10, ER14, EUECTYP5, EUECTYP6, T05AMT, T06AMT, T10AMT, T14AMT, EDISABL, EDISPREV)
5. Health Insurance (ECDMNTH, ECRMNTH, EHEMPY, EHIMTH, EHIOWNER, EMCOCOV, RCHAMPM, RMEDCODE, RPRVHI, RPRVHI2, RCUOW58A, RCUOW58B, RCUTYP58)
6. Social Security (RCUOWN01, RCUTYP01, ER01A, ER01K, T01AMTA, T01AMTK, ECRMTH, RMEDCODE, TFSOCSEC)
7. Social Welfare Programs (RCUOWN03, RCUOWN04, RCUOWN25, RCUOWN27, RCUTYP03, RCUTYP04, RCUTYP25, RCUTYP27, TFSSI, TFTRNINC, EFSYN, EWICYN, EPATANF1, EPATANF2, EPATANF3, EPATANF4, EPATANF5, EPATANF6, ER03A, ER03K, ER04, T03AMTA, T03AMTK, T04AMT)
8. Asset Ownership (EAST2D, EAST1B, EAST2A, EGVJT, ECDJT, ECKJT, EMDJT, EBDJT, ESVJT, EAST2C, EAST3E, EMRTJNT, EMRTOWN, EAST3A, EAST3C, EAST4C, ESVOAST, EAST4A, EAST4B, EGVOAST, ECDOAST, ECKOAST, EMDOAST, EBDOAST, EAST3B, EAST3D, EAST1A)

The recording of these events will be evaluated based at multiple levels of agreement. Using unweighted distributions of reporting for the same time period each household with events recorded via both SIPP 2004 and the EHC test will be evaluated for consistency on a month-by-month basis for each domain (1). Reports recorded in a monthly two-by-two table will show consistency and over/underreporting.

(1) 2x2 Report Consistency Tables

SIPP/EHC Report Consistency in <u>[MONTH]</u> for <u>[PROGRAM/CHARACTERISTIC]</u>			
		SIPP	
		yes	no
EHC	yes	a	b
	no	c	d

outcomes: if $b = c$: equivalent data quality
 if $b > c$: “underreporting” in SIPP, relative to EHC
 if $b < c$: “underreporting” in EHC, relative to SIPP

As outlined above, the data quality analyses will focus on the measurement and repeat measurement of respondents’ events in both the SIPP and the EHC for calendar year 2007. By generating these tables for each month of the reference period we will be able to determine if the level of underreporting in the EHC is greater than in the SIPP for the first and second third of the reference period relative to the last third (the longer recall periods where the EHC is expected to suffer relative to the SIPP). In addition to these two-by-two tables measuring exact monthly correspondence, we will also broaden the agreement to +/- 1 month from an exact match and examine the occurrence and timing of events with the following categorical variables. These will be constructed by domain and examined for the whole time period as well as for sections of the reference period. The percent distribution by domain in these outcome variables will be evaluated to determine where differences occur and in which direction (greater or lesser reporting of events in SIPP versus the EHC).

OCCUR (All cases)

1. Spell in both SIPP and EHC
2. Spell in SIPP not EHC
3. Spell in EHC not SIPP
4. No spell in SIPP or EHC

TIMING (Cases with spells in both SIPP and EHC)

1. SIPP and EHC agreement on month
2. SIPP and EHC 1 month difference in incidence month
3. SIPP and EHC 2-4 months difference in incidence month
4. Spell in both SIPP and EHC more than 4 months difference in timing

Other data quality differences may be suggested by the quality of the distributions of spell transitions across calendar months. This phase of the analysis will compare the levels and patterns recorded in each of the three interview components: SIPP 2004, EHC (primed), and EHC (un-primed). Comparison of the data recorded from the two groups of EHC respondents will provide a way to examine the effect of priming introduced due to the re-interview design. There will be respondents in the un-primed group that will have some baseline data – allowing background patterns of program receipt to be used in the evaluation of this group’s data as well.

We will focus analyses for each domain on the relative timing during the calendar year of events. This will allow us to address concerns that the reporting of events degrades with a longer recall period. As described, the EHC is a tool to aid in recall and improve consistency over topical domains. If successful the EHC will not substantially underreport events at the beginning of the year relative to the reporting of events at the middle or end of the year. To evaluate this, we will be considering the distributions of events over the thirds of the year. Due to the rotational nature of the SIPP sample these thirds will not easily overlay the waves in the SIPP, but SIPP events and distributions can be no more than 4 months from the interview, and will still provide a good comparison even though the each third of the EHC reference year will overlap waves and reference months in the SIPP data (see Figure 1 above).

The second avenue of comparison will be to evaluate report consistency for the total calendar months of participation or coverage in each topic area. To evaluate the duration and prevalence we will compare 13x13 tables (2) in which rows correspond to the number of months in calendar 2007 with a “yes” on a variable such as unemployment insurance or social security receipt in the EHC, columns correspond to the same measure from the SIPP 2004 panel data, and cells in the table contain the number of interviews falling in that (months in SIPP, months in EHC) cell. If EHC reproduces SIPP perfectly, only the diagonal cells will have non-zero entries. If it does not, such a table may help to illuminate the patterns of bias. As with the previously described tables we plan to disaggregate these distributional comparisons (percent TANF, percent food stamps, etc.) by 4-month period, as another means of examining possible biases caused by the longer EHC recall period.

We are careful not to identify our results in terms of better or worse for most of these comparisons. Events that occur in SIPP on seams may be erroneous; certainly the timing of these events is suspect if they are concentrated on seams. For most of the analyses we will focus on unedited SIPP 2004 data for comparability with the EHC data, which also will be unedited.

Seam and non-seam transitions will be evaluated as well. With the analysis table (3) below we examine the transition rates for seam months, separately from non-seam months and compare both with transitions observed in data collected with the EHC.

(3) Month-to-Month Transition Rates in 2007 (Selected Calendar Month-Pairs Waves 10/11/12) among SIPP Seam Cases, SIPP Off-Seam Cases, and EHC Cases.

CY 2007 Month-Pair	Interviewed Rotations	Seam Cases	Off-Seam Cases
JAN-FEB	all	r1 (w10/11)	r2 (w10), r3 (w10), r4 (w10)
FEB-MAR	all	r2 (w10/11)	r1 (w11), r3 (w10), r4 (w10)
MAR-APR	all	r3 (w10/11)	r1 (w11), r2 (w11), r4 (w10)
APR-MAY	all	r4 (w10/11)	r1 (w11), r2 (w11), r3 (w11)
MAY-JUN	all	r1 (w11/12)	r2 (w11), r3 (w11), r4 (w11)
JUN-JUL	all	r2 (w11/12)	r1 (w12), r3 (w11), r4 (w11)
JUL-AUG	all	r3 (w11/12)	r1 (w12), r2 (w12), r4 (w11)
AUG-SEP	all	r4 (w11/12)	r1 (w12), r2 (w12), r3 (w12)
(Note: The final three month-pairs of 2007 are not available for this analysis.)			
SEP-OCT	r2, r3, r4	–	r2 (w12), r3 (w12), r4 (w12)
OCT-NOV	r3, r4	–	r3 (w12), r4 (w12)
NOV-DEC	r4	–	r4 (w12)

Month-to-Month Transition Rates in 2007 (Selected Calendar Month-Pairs) among SIPP Seam Cases, SIPP Off-Seam Cases, and EHC Cases for <u>[program/characteristic]</u>								
	Jan- Feb	Feb- Mar	Mar- Apr	Apr- May	May- Jun	Jun- Jul	Jul- Aug	Aug- Sep
SIPP - Seam Cases								
SIPP - Off-Seam Cases								
EHC								

An additional source for potentially erroneous transitions is the SIPP edit process. Many topic areas are edited within the wave they are collected without respect for the prior reports. Changes in respondent or missing waves of data can generate transitions separately from those reported, often these occur on the seam but in many cases the edit process assigns them to months during

the reference period. This is another reason that we will focus on the unedited data when comparing the occurrence of transitions between these two data collection instruments.

Distributional characteristics, such as the percent with TANF, Food Stamps, Medicare, Working, Enrolled, and with Health Insurance coverage from the EHC will be compared to the same distributions from SIPP. This component of the analysis will begin to inform us about the work that will be necessary in bridging estimates produced via the two data collection systems. We will produce indices of dissimilarity, indicating how much one distribution would have to be adjusted to mirror the other. These distributional comparisons will also be done for the different portions of the reference year separately to add to the evaluation of the possible degradation in recall for the early portions of the year in the EHC relative to the SIPP.

The inter-domain consistency will be evaluated to determine the relative timing of events across topics. We expect the EHC will significantly improve the consistency across domains, and this will be analyzed by looking at the correlations between events from different topic domains in both SIPP and EHC and see which has stronger correlations. Are simultaneous changes across domains reported consistently in both instruments? The occurrence of simultaneous events will be evaluated with correlations based on pair-wise comparisons and on factor analysis of events in different domains.

The validation component of the analysis, where we compare both SIPP 2004 responses and EHC responses to administrative records depends on reaching the necessary data agreements with the administrative data sources. The first component of the analysis, re-interview and comparison of SIPP and EHC data can proceed before the data agreements are finalized. Substantial groundwork has already been laid to be able to utilize administrative records for several programs (e.g., TANF, Food Stamps, Medicare, Social Security, SSI, and possibly wage information). This validation stage of the analysis will occur after the first stage comparisons due to the added time necessary to obtain and match the necessary administrative records. Once administrative records are available, programs with comparison data will be added to the distributional comparisons described above where we can generate them from the records in the validation component of the analysis. The validation analyses will be key in determining accuracy among the reported programs. As with any validation exercise, there is the possibility that a respondent will report receiving benefit or being on a program with no record in the administrative record – this can occur due to respondent confusion about which program they receive or lags in the records system. There is also the possibility that a respondent would be identified as receiving a benefit in the administrative records but fail to report it in the survey instrument, most likely a situation exemplified by recall error – we would expect this type of error to be more prevalent in the early portion of the EHC recall period to a greater degree than in the later portion or in the SIPP.

Additional evaluation methods – respondent debriefings, interviewer debriefings and focus groups, interview observations, analysis of recorded interviews, etc. – will be directed toward a

better understanding of the EHC interview process, such as how landmark dates are introduced and used, the preferred “direction” of reporting, the extent to which events in one domain are used to pinpoint transitions in another domain, etc.

By including direct comparisons across survey instruments, as well as an administrative-record-based validation component, this research will be able to add significantly to the literature on event history calendar survey methodology, especially with respect to validating the SIPP and EHC reporting of income transfer program receipt and amounts over a calendar year. Results from the study will also inform the decision of whether to use EHC methods in the re-engineered SIPP program currently under development at the Census Bureau.

Next Steps

Following the 2008 paper instrument evaluation, (assuming a positive outcome) a broad dress-rehearsal evaluation of the new electronic EHC instrument being designed for the re-engineered SIPP for possible administration in September 2009. The results from the 2008 EHC evaluation will be used to refine training procedures and make necessary adjustments to the new computer assisted personal interview (CAPI) EHC being prepared for the dress rehearsal.

The planning and instrument development for the 2009 re-engineered SIPP dress rehearsal is well underway. The survey is scheduled to be administered in September – the earliest possible administration window for the dress rehearsal. It will collect information about jobs, programs, health insurance and demographics for the 2008 calendar year. The dress rehearsal will implement the lessons learned in developing field procedures for the 2008 EHC evaluation and extend field implementation to each of the Regional Offices for this national test. The 2009 dress rehearsal instrument will be evaluated in several domains including field implementation issues and data comparability vis-à-vis SIPP 2008 and administrative records. The administration of the 2009 dress rehearsal in September is not ideal, but is the earliest in 2009 that the instrument can be ready for implementation. The production implementation of an EHC in the re-engineered SIPP would be during the early part of the calendar year to minimize the length of recall in the reporting of data for the prior calendar year. Results from both the 2008 evaluation and the 2009 dress rehearsal will be used to make final decisions regarding the design and implementation of the re-engineered SIPP for production in 2011 or 2012.

References

- Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar. Potential improvements in the quality of retrospective reports in surveys. *Memory*, 6, 383-406.
- Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly*, 65, 45-74.
- Callegaro, M. (2007). *Seam Effects Changes Due to Modifications in Question Wording and Data Collection Strategies, A Comparison of Conventional Questionnaire and Event History Calendar Seam Effects in the PSID*. Lincoln, NE: University of Nebraska
- Cantor, D. and Levin, K. (1991), "Summary of Activities to Evaluate the Dependent Interviewing Procedure of the Current Population Survey," Westat, Inc.: report submitted to the Bureau of Labor Statistics (Contract No. J-9-J-8-0083).
- Caspi, A., Moffitt, T. E., Arland, T., Freedman, D., Amell, J. W., Harrington, H., et al. (1996). The life history calendar: A research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research*, 6, 101-114.
- Collins, C. (1975), "Comparison of Month-to-Month Changes in Industry and Occupation Codes with Respondent's Report of Change: CPS Job Mobility Study," U.S. Census Bureau, Response Research Staff Report No. 75-5, May 15, 1975.
- Ensel, W. M., Peek, K. M., Lin, N., & Lai, G. W.-f. (1996). Stress in life course. A life history approach. *Journal of Aging and Health*, 8, 389-416.
- Fields, J. and Callegaro, M. (2007) Background and Planning for Incorporating an Event History Calendar into the Re-Engineered SIPP. Prepared for presentation at the Federal Committee for Statistical Methodology conference Nov. 2007.
- Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: a technique for collecting retrospective data. *Sociological Methodology*, 18, 37-68.
- Hill, D. (1994), "The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey," *Journal of Official Statistics* 10: 359-380.

- Hoogendoorn, A. (2004), “A Questionnaire Design for Dependent Interviewing that Addresses the Problem of Cognitive Satisficing,” *Journal of Official Statistics* 20: 219-232.
- Kalton, G. and Miller, M. (1991), “The Seam Effect with Social Security Income in the Survey of Income and Program Participation,” *Journal of Official Statistics* 7: 235-245.
- Kominski, R. (1990). The SIPP Event History Calendar: Aiding respondents in the dating of longitudinal events. In Proceedings of the Section of Survey Research Methods (pp. 553-558). Washington D.C.: American Statistical Association.
- LeMaître, G. (1992), “Dealing with the Seam Problem for the Survey of Labour and Income Dynamics,” Statistics Canada: SLID Research Paper Series, Catalogue No. 92-05, August 1992.
- Lynn, P. and Sala, E. (2006), “Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing,” *International Journal of Public Opinion Research* 18: 500-509.
- Martini, A. (1989), “Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP,” Proceedings of the American Statistical Association, Section on Survey Research Methods, 387-392.
- Michaud, S., Dolson, D., Adams, D., and Renaud, M. (1995), “Combining Administrative and Survey Data to Reduce Respondent Burden in Longitudinal Surveys,” Statistics Canada: SLID Research Paper Series, Catalogue No. 95-19 (Product Registration Number 75F0002M), August 1995.
- Moore, J. and Marquis, K. (1989), “Using Administrative Record Data to Evaluate the Quality of Survey Estimates,” Survey Methodology 15: 129-143.
- Polivka, A. and Rothgeb, J. (1993), “Redesigning the CPS Questionnaire,” Monthly Labor Review, September 1993, 10-28.
- Rips, L., Conrad, F., and Fricker, S. (2003), “Straightening the Seam Effect in Panel Surveys,” Public Opinion Quarterly, 67: 522-554.
- Stanley, J. and Safer, M. (1997), “‘Last Time You Had 78, How Many Do You Have Now?’ The Effect of Providing Previous Reports on Current Reports of Cattle Inventories,” Proceedings of the American Statistical Association, Section on Survey Research Methods, 875-880.

Yoshihama, M., Gillespie, B., Hammok, A. C., Belli, R. F., & Tolman, R. M. (2005). Does the life history calendar method facilitate the recall of intimate partner violence? Comparison of two methods of data collection. *Social Work Research*, 29, 151-163.

Young, N. (1989), “Wave-Seam Effects in the SIPP,” Proceedings of the American Statistical Association, Section on Survey Research Methods, 393-398.