

RESEARCH REPORT SERIES
(*Survey Methodology #2009-09*)

**The 2008 Survey of Income and Program Participation
Event History Calendar Field Test:
Study Design and Initial Results**

Jeffrey Moore
Jason Fields¹
Gary Benedetto¹
Martha Stinson¹
Anna Chan
Jerry Maples

¹ Housing and Household Economic Statistics Division

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: December 30, 2009

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

The 2008 Survey of Income and Program Participation Event History Calendar Field Test: Study Design and Initial Results[#]

Jeffrey Moore⁽¹⁾, Jason Fields⁽²⁾, Gary Benedetto⁽²⁾,
Martha Stinson⁽²⁾, Anna Chan⁽¹⁾, and Jerry Maples⁽¹⁾

⁽¹⁾Statistical Research Division

⁽²⁾Housing and Household Economic Statistics Division
U.S. Census Bureau

Paper prepared for presentation at the annual meetings of
the American Association for Public Opinion Research
Hollywood, FL, May 14-17, 2009

ABSTRACT: The U.S. Census Bureau's Survey of Income and Program Participation (SIPP) provides monthly information about the nation's income, wealth, and program usage. Currently, SIPP administers three interviews per year to each sample member; each interview's reference period covers the preceding four calendar months. In 2006 the Census Bureau initiated a SIPP re-engineering effort, a key component of which is a shift to a single annual interview covering the preceding calendar year. To accomplish this shift, the Census Bureau proposes to employ event history calendar (EHC) methods. Prior research, however, has raised some questions about EHC data quality for topics of key importance to SIPP, such as need-based program participation. In addition, the research base does not address the main SIPP design issue – the proposed shift from a four-month to a twelve-month reference period. To examine the implications of the switch to EHC methods and the corresponding expansion of the survey's reference period, the Census Bureau conducted a field test of a prototype paper version of an EHC-based SIPP instrument test in the spring of 2008. The essential feature of the test was an EHC reinterview of expired SIPP 2004 panel households. The reference period for the reinterview was calendar year 2007; the primary sample component consisted of cases which had already provided information about calendar year 2007 in the course of their final three SIPP interviews. The field test thus permits a direct comparison of standard questionnaire and EHC reports by the same people, about the same characteristics, and for the same time period. This paper documents the design of the 2008 "paper" test and describes some initial results with regard to the correspondence of the two survey reports – one obtained from a standard questionnaire, the other from an EHC instrument – for several of the key characteristics of interest to SIPP, and for each month of 2007.

Key words: questionnaire design, recall, reference period, reinterview, response error

[#] This paper draws on and extends a previous paper by Moore and Fields (2008). It is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are the authors' and not necessarily those of the U.S. Census Bureau. We offer special thanks to Joanne Pascale, who has been a valuable colleague on this research project from the beginning, and who provided many useful comments on an earlier draft of this paper. We also thank our colleagues Judy Eargle, Don Malec, and Chuck Nelson for their careful reviews and their thoughtful comments and suggestions.

The 2008 Survey of Income and Program Participation Event History Calendar Field Test: Study Design and Initial Results

Jeffrey Moore, Jason Fields, Gary Benedetto, Martha Stinson, Anna Chan, and Jerry Maples
U.S. Census Bureau

1. Introduction and Overview

This paper describes and presents initial results from a field test conducted in 2008 to evaluate the use of new interviewing methods in a major longitudinal survey program of the U.S. Census Bureau – the Survey of Income and Program Participation (SIPP). The new methods consist primarily of an annual interview (as opposed to multiple interviews throughout the year) using event history calendar (EHC) methods to capture key data. The 2008 field test – also called the “paper” test – was in essence a re-interview of former SIPP respondents using a prototype, paper-and-pencil EHC instrument, thus allowing a comparison of responses from the same people, for the same time period (calendar year 2007), and concerning the same characteristics, the key difference being whether those responses were obtained from a series of 4-month reference period interviews (SIPP) or from a single EHC interview covering the entire calendar year. This paper focuses on the initial evaluation results comparing the two survey reports. Subsequent work will address additional comparisons – e.g., month-to-month transitions and income amount reports – and will also include administrative record data, thus permitting a clearer and more direct comparison of data quality for selected characteristics.

The 2008 “paper” test was the first in a planned series of research efforts to demonstrate the feasibility of an annual, 12-month reference period SIPP interview, using EHC methods, and to evaluate the quality of the resulting data. The test was viewed as providing a “go” or “no-go” signal for further testing and development. The conclusion: go. Planning and implementation work is currently underway for a large-scale test of a fully automated instrument incorporating EHC procedures in early 2010. Currently, the production version of the re-engineered SIPP program is scheduled for launch in 2013.

2. Background

2.1. The Survey of Income and Program Participation (SIPP)

SIPP is a nationally-representative, interviewer-administered, longitudinal survey conducted by the U.S. Census Bureau. Since its 1984 inception, SIPP’s main objective 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 annual and sub-annual income dynamics, movements into and out of government transfer programs, family and social context of individuals and households, and the interactions among these items. 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.

In its current design, each SIPP panel consists of multiple waves (or rounds) of interviewing,

with waves administered three times a year, at four month intervals. The SIPP sample is split into four equivalent subsamples, called “rotation groups”; each rotation group’s interview schedule is staggered by one month, in order to maintain a constant workload for field staff. Starting with the 1996 panel, all SIPP interviews are now conducted with a computer-assisted questionnaire. The first interview is administered in-person; subsequent interviews are generally conducted via telephone. The SIPP core instrument, which contains the survey content that is repeated in every survey wave, is detailed, long, and complex, collecting information about household structure, labor force participation, income sources and amounts, educational attainment, school enrollment, and health insurance over the prior four-month period. Various “topical modules” appended to the core once per year collect detailed annual data on taxes, assets, and liabilities; additional modules collect data once per panel on a variety of other topics, including respondents’ marriage, fertility, employment, and program benefit receipt histories. A typical SIPP interview takes about 30 minutes per interviewed adult. See U.S. Census Bureau (2001a) for a more complete description of the current SIPP program.

2.2. The Re-Engineering of SIPP

In 2006 the Census Bureau initiated efforts to re-engineer the SIPP program. Perhaps not surprisingly, the main catalyst for the re-engineering was a budget crisis. In early 2006, the Census Bureau was faced with a projected shortfall of approximately 40 million dollars in its fiscal year 2007 budget. The agency’s lesser-of-two-evils response was to place the entire burden of the shortfall on a single program, SIPP, rather than spreading it out and burdening many programs with incomplete funding. Although there were some signs of commitment to revising the program in a new form at some later time, the very existence of the SIPP program was clearly threatened, and at a minimum it appeared that there would be a break in the data series of several years’ duration. Policy-makers, data users, and other stakeholders protested strongly to Congress, emphasizing the unique value of the survey. They decried the proposed “data gap,” and also expressed grave concern that the gap might become a permanent victim of other funding priorities.

Ultimately, Congress opted to continue the survey, providing funding for both continuing the current SIPP design and, simultaneously, developing a broad array of improvements for the future – the “re-engineering” effort. Those improvements, to be implemented in a future panel, and without harming data quality, include (1) reduction of costs; (2) reduction of respondent burden; (3) improvements in the processing system; (4) improvements in the collection instrument – more specifically, the development and evaluation of a Windows-based instrument authored in the BLAISE language, as required by current laptop technology, and the use of EHC methods in the collection of a calendar year’s worth of month-level data; (5) increased use of administrative record data, as both a supplement to the survey data and a means of evaluating their quality; and (6) development of survey content and procedures in close collaboration with data users and other stakeholders.

A key component of the re-engineered SIPP is a proposed shift away from its traditional three-times-per-year data collection schedule, with each interview covering the preceding four calendar months, to a single annual interview covering the preceding calendar year. Such a change would yield several key benefits almost immediately – field costs would be substantially reduced, for example, as would respondent burden, and the SIPP data structure, freed from its

traditional staggered “rotation group” design, would be substantially simplified. To maintain data quality while accomplishing this shift the Census Bureau proposes to employ event history calendar (EHC) methods to gather SIPP data (Fields and Callegaro, 2007).

These prospective design changes are major ones. In order to ensure that the resulting data meet the needs of SIPP users, they require that the Census Bureau devote substantial effort and resources to developing, testing, and refining the re-engineered survey. The 2008 “paper” test (the focus of this paper) is the first stage of that program. Assuming sufficiently positive outcomes from the 2008 test, a larger-scale test of a fully automated instrument will take place in early 2010. Evaluation of the data from the 2010 test will indicate areas of needed refinement for the future production version of the survey. At this point the Census Bureau expects the re-engineered SIPP to be fully operational starting in 2013, collecting information covering calendar year 2012. Continuation of the 2008 panel interviews through the end of 2012 will permit some “benchmarking” of comparable estimates from each source for the same time period, which will help shed light on the cause of any discontinuities in the data series and the appropriate steps to ameliorate them.

2.3. Event History Calendar Interviewing

EHC methods have been employed since the 1960’s to assist interviewers in collecting detailed data across long recall periods (Belli, 1998; Belli, Shay, and Stafford, 2001; Callegaro 2007). The essential and unique feature of EHC methods is their approach to assisting recall of information through the exploitation of naturally-occurring links and associations between and among memory elements. Each to-be-recalled spell or event can happen before, after, or at the same time as another spell or event; the sequence may have a causal basis, or it may represent simple happenstance. An example of the former would be a change in employment which necessitates a residence change (or vice-versa); an example of the latter would be some sequence of a family event (e.g., a wedding, a vacation) and a meteorological one (a hurricane, a blizzard). 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-list or question-by-question (Q-by-Q) approach.

The survey methods research literature documents several evaluations of EHC instruments, most of which involve a comparison of EHC results with data derived from a standard, Q-by-Q questionnaire administered previously to the same set of respondents (e.g., Freedman et al. 1988; Caspi et al. 1996; Ensel et al. 1996). Belli, Shay, and Stafford (2001) offer a slight variation – a test-retest experiment wherein prior respondents were reinterviewed by two treatment groups, one with a two year EHC and one with a traditional Q-by-Q instrument. The general findings from all of the studies comparing EHC and Q-by-Q methods is that, across most domains, EHC-derived data are at least as good as, and often better than, data obtained from standard Q-by-Q questionnaires, in terms of agreement between the later data collection using EHC methods and the original data collected earlier. Yoshihama and colleagues (2005), in their study of intimate partner violence reporting, arrived at a similar conclusion using a very different approach. They compared reporting of violence incidents in two samples of women, one interviewed with traditional interviewing and one with a life history calendar. The EHC approach yielded significantly more reported incidents of intimate partner violence, which the authors conclude is an indication of a more effective interviewing technique, given the highly sensitive nature of the

target events. A recent volume by Belli, Stafford, and Alwin (2008) provides an excellent general summary of EHC methods and the current research literature concerning data quality.

An important exception to the general conclusion of equivalent or better data quality is the quality of EHC data for need-based government programs. For example, while Belli and colleagues (2001), in their Panel Study of Income Dynamics (PSID) research, find better EHC recall for most topics, this was not the case for reporting on AFDC and Food Stamps receipt, where no significant difference was found. Many caveats are in order: the research lacked a strong and objective measure of quality; the EHC approach did not perform *worse* than the standard Q-by-Q questionnaire, it simply did not perform better; these characteristics were not the primary focus of the survey or of training; there is no immediately obvious reason why recall of these programs should differ from recall of other types of events; there is little if any other evidence in the literature of a problem with EHC data collection for these characteristics, etc. Nevertheless, the null finding in the PSID study is somewhat troubling for SIPP, and one of the primary motivators for the field test research described here.

In addition, the research base is quite limited in terms of a direct bearing on the main SIPP design issue – the proposed shift from a reference period which covers the four preceding calendar months to one which covers the preceding calendar year. Almost exclusively, published studies comparing the Q-by-Q and EHC interview treatments have used survey recall periods much longer than SIPP’s – none less than a year, and some covering multiple years. Thus, concerns linger about the data quality implications of both the switch to EHC methods for the topics covered in SIPP, as well as the shift from a four-month recall period to a one-year recall period.

2.4. Census Bureau Experience with EHC Methods

Although the Census Bureau has never implemented EHC interviewing in a production survey, the agency does have experience with closely related methods. In the late 1980’s an EHC-like interviewing aid was field tested in SIPP in a limited geographic area (Kominski, 1990). Although the test was judged a success, the likely gains did not outweigh the perceived costs of integrating the calendar aid into the data processing system, and thus its implementation as a standard component of the SIPP interview was judged to be impractical. Only in the last decade or so have electronic EHC instruments been developed and refined, significantly easing some of the issues associated with retrieving and coding the data collected with this tool.

3. 2008 EHC Field Test Design

This section describes the general design of the initial “feasibility” test of the use of EHC interviewing methods for the re-engineered SIPP, consisting of a relatively small-scale field test of a prototype EHC questionnaire covering calendar year 2007, administered in early 2008 to expired 2004 panel SIPP households. The sample component of primary interest consists of cases which were interviewed through the end of the 2004 panel, and who thus reported about calendar year 2007 (January through September, at a minimum) via their wave 10, 11, and 12 interviews – the final three waves of the SIPP 2004 panel; see Figure 1¹. This sample permits a

¹ All figures may be found in Appendix 1.

direct comparison of two survey reports produced by the same people and for the same months – one obtained from a standard SIPP questionnaire, and the other from a 12-month questionnaire employing EHC methods. Although not included in this report, administrative record data will eventually be made available to the project for a small subset of the characteristics measured in both questionnaires, thus enabling an objective assessment of data quality differences for those characteristics.

A second sample component in the 2008 test is comprised of cases interviewed only through wave 8, after which they were cut from the 2004 panel sample. We use these cases to address one of the important limitations of the test design, the possibility that the results might be affected by some form of “priming bias” – see section 3.2.

In an effort to minimize nonresponse, the field test included a \$40 incentive for all sample households, which interviewers were to offer independent of the household’s survey cooperation. The incentive came in the form of a debit card promised at the time of the interviewer’s visit, and mailed to the household from the coordinating Regional Office within two weeks of that visit.

3.1. The Field Test Instrument

The field test employed a paper questionnaire, administered in person, covering a small subset of subject-matter areas of interest to SIPP²; time and resource constraints did not allow for the development of a computer-assisted questionnaire. While this is certainly a limitation of the study design (see section 4), we did not consider the abbreviated paper questionnaire to be a serious impediment to conducting the research, since we viewed the test as an evaluation of the feasibility of EHC methods in general, as opposed to an evaluation of any specific rendering of those methods. We contracted with RTI International to develop and refine a draft instrument and accompanying materials, and an interviewer training package, and to pilot test them with a small group of interviewers, using a small, “convenience sample” of respondents. This contract proved very useful; even though we made substantial modifications to the pilot test materials, the drafts produced by RTI provided the foundation for the 2008 field test instrument and training package.

As noted above, we selected a limited set of key subject-matter topics (“domains”) for inclusion on the field test instrument. These included residences, school enrollment, labor force participation, workers’ insurance programs (unemployment, disability, and workers’ compensation), Social Security, the major government programs which provide “welfare”-type assistance to the poor (Temporary Assistance to Needy Families (TANF), Food Stamps, the Women, Infants, and Children (WIC) nutrition program, and Supplemental Security Income (SSI)), health insurance coverage, and asset ownership. The asset ownership domain differed from the others in that it was designed to capture simply the fact of ownership, at any time during the calendar year 2007 reference period, of any of several types of retirement accounts and asset types. All of the other domains attempted to capture receipt/participation events and spells at the month level (the third-of-a-month level for employment-related characteristics).

Even though the true period of interest was calendar year 2007, the EHC questionnaire

² A copy of the EHC questionnaire is available from the first author upon request.

approached almost all domains by starting with a “now” or “currently” question – “Are you enrolled in school now?” “Are you currently receiving unemployment benefits?” (In fact, we employed this approach for all of the domains for which results are included in this report.) This does not appear to be standard practice with EHC interviewing (for example, see Belli, Stafford, and Alwin, 2008). We opted for this approach based on two assumptions: (1) A “now/currently” question is perhaps the easiest possible question for a respondent to answer. And (2), a “yes” response yields a simple and natural and concrete starting point for talking about the target reference period – when did that spell start?; has it been continuous since then?; (if appropriate) were there other months in 2007?; etc. – and for using the particular strengths of the EHC method to assist the respondent with his/her recall task. If the “now/currently” question elicited a “no” response, then the follow-up moved directly to the reference period – “How about at any time during 2007?” – with additional probes about specific months, as appropriate³. In contrast, SIPP’s standard approach to its reference period is to start by asking about the entire period (plus the interview month up to the date of the interview):

“At any time since [month1] 1st, did you receive benefits from WIC – the Women, Infants, and Children nutrition program?”

and then, if yes, to follow-up with questions about each month:

“Earlier I recorded that you received WIC, or the Women, Infants, and Children nutrition program. Have you received any benefits from the WIC program yet this month? In which of the last four months – [month1], [month2], [month3], or [month4] – did you (.../also) receive WIC benefits?”

In one important sense the response task in the EHC interview was considerably more difficult than in the SIPP interview, at least for some respondents. SIPP employs dependent questions which, after the initial interview, offer respondents reminders of their status at the end of the previous wave’s reference period before asking about the current wave. For example, a respondent who reports Food Stamps receipt in the last month of the wave n reference period is asked a question of the following sort in wave $n+1$: “Last time I recorded that you received Food Stamps in [month]. Did you continue to receive Food Stamps after that?” Moore (2008) provides a more complete description of SIPP’s dependent procedures, and presents strong evidence of their substantial data quality benefits. Dependent procedures were not part of the EHC interview design for several reasons, including the difficulty of implementing dependent procedures in a paper-and-pencil questionnaire, the added time and burden of identifying who was and who was not a former SIPP respondent at the outset of the EHC interview, and the added complexity of having to adjust the dependent/independent nature of the questions based on that information; and the confounding effect dependent procedures would have on the priming bias analysis (see section 3.2).

3.2. The Field Test Sample

Several factors were paramount in the selection of sample cases for the field test: resource

³ The “now/currently” question was explicitly scripted on the EHC questionnaire. The follow-up questions, however, were in the form of very abbreviated cues to the interviewer. For example, this is what was printed on the form following a “yes” response to the question: “Do you receive WIC benefits now?”:

[When start? Continuous? Any other times in 2007? When/What months?]

The interviewer cues following a “no” response were similarly abbreviated:

[Any time in 2007?] []-Yes []-No [When/What months?]

constraints, of course, both in terms of funding and with regard to the availability of field staff; the eventual availability of administrative record data – in particular, data for state-administered programs such as TANF and Food Stamps – for a more objective assessment of data quality differences; and the number and concentration of SIPP sample cases in states where administrative record data access agreements were either in place or where the prospects for such agreements were good. This combination of factors eventually led us to select Illinois and four metropolitan areas in Texas – Austin, Dallas/Fort Worth, Houston, and San Antonio – as the sites of the EHC field test. The primary sample consisted of all SIPP 2004 panel cases in those areas which had completed a wave 11 interview.

Concerns about one potential source of bias led to the selection of a second major sample component. An obvious limitation of the research design is the possibility for some form of “priming” effect – specifically, that the responses of EHC respondents about calendar year 2007 events and circumstances might be affected by their already having reported about those same events and circumstances via their SIPP interviews. Serendipitously, the 2004 SIPP panel provided us with a parallel sample of cases which had not been so primed. Due to an earlier budget shortfall, SIPP made a substantial sample cut following the wave 8 interview (interviews in June through September of 2006), after which a systematic sample of cases received no more SIPP interviews – see Flanagan (2006) for details. The wave-8-sample-cut cases in Illinois and the four Texas metropolitan areas were also included in the EHC field test. The extent to which the distributions of the EHC reports from these cases are in accord with those of the “continuing” wave 11 cases will provide important information about the impact of priming on the latter group, which will in turn affect the nature of and the level of confidence in the conclusions that can be drawn from the primary comparison of interest – Q-by-Q and EHC reports for the same time period⁴.

The initial sample size for the EHC field test was 1,945 addresses, consisting of 1,096 continuing wave 11 cases and 849 wave-8-sample-cut cases, distributed by state as shown in Table 1⁵. Note that interviewers’ assignments consisted simply of addresses at which to conduct the EHC interviews, not specific people or households. That is, they were not provided any advance information about the expected residents of each sample address, and their instructions were simply to interview whomever they found. An after-the-fact clerical procedure matched EHC field test households and people to SIPP records to identify those who had, in fact, participated in SIPP, either via a self-response or a proxy interview, versus those who had not. The EHC field test cases of primary analytical interest for this paper are those from the continuing wave 11 subsample which were successfully matched to SIPP. These are the cases which provide calendar year 2007 data from both their SIPP interviews and from the EHC field test.

3.3. Administrative Records

For a subset of characteristics, administrative record data will provide a rigorous and independent means of assessing and comparing the quality of the data captured through the EHC

⁴ Because they were no doubt quite familiar with the type of information of interest to SIPP, the wave-8-sample-cut cases cannot be said to address all types of potential “priming” effects, only those having to do with already having provided data about calendar year 2007.

⁵ All tables referred to in the body of this paper may be found in Appendix 2.

and the standard SIPP interview methods. All use of such data, however, is strictly contingent on having gained respondents' consent to do so. Starting with wave 8 of the 2004 SIPP panel, all respondents were given an explicit opportunity to opt out of any procedures by which their SIPP reports would be linked to administrative records; those who opted out will be excluded from the match to administrative record data. Eventually, the 2008 field test will have access to the following records covering calendar year 2007: Medicare, OASDI ("Social Security") retirement and disability benefits, SSI, public housing, and quarterly employment status and earnings (at the federal level); and TANF and Food Stamps (at the state level). Evaluation results using administrative records to assess data quality will be the subject of a future report.

3.4. Field Test Interviewers and Interviewer Training

A total of 107 Census Bureau field representatives (FRs) working from the Chicago and Dallas Regional Offices conducted the EHC field test interviews. Most had at least some experience interviewing for other Census Bureau surveys or for SIPP; a few were new hires, brought on board specifically to staff the EHC project (see Table 2). None of the interviewers had any prior experience with EHC interviewing procedures.

EHC interviewing requires a new set of skills. An effective EHC interviewer cannot just read the questionnaire script, but must remain keenly alert to signs of data quality trouble, and must know what tools to call upon in the face of that trouble. He or she must be able to (a) recognize when a respondent is having difficulty reporting accurately (or, apart from any obvious difficulty, to recognize when recalled material is illogical, or inconsistent, or otherwise "not right" somehow); (b) in the face of many competing demands on attention, to remain open to the receipt of distress signals, and (c) must have an available response repertoire of effective techniques to call upon to help the respondent achieve better quality recall, drawing upon the links between and among memory elements. These skills are not easy to acquire, perhaps especially in an interviewing culture which primarily emphasizes and values productivity, minimal nonresponse, and reading questions as worded.

Field test FRs received three days of training on EHC concepts and practices; an additional half day was devoted to administrative matters, and another half day, for new hires, focused on general interviewing practices (preparation, planning an efficient travel route, rules of behavior, maximizing respondent cooperation, etc.)⁶. As with the instrument itself, the training package built upon and benefited from the initial developmental work of our RTI contractors, which included a successful pretest of a draft training package and the initial EHC questionnaire and other materials. However, the major modifications made to the final EHC form (noted earlier) necessitated major modifications to the training materials as well. One key aspect of the training that we did maintain – and even extended somewhat – was a heavy emphasis on very active learning, as opposed to lecturing, including many demonstrations and paired-practice interviews. It is unclear whether three days was sufficient for inculcating a very different outlook on interviewing and very different interviewing practices, but both trainees and trainers expressed high praise for the EHC training package in their post-training comments and in debriefing sessions held at the conclusion of interviewing.

⁶ A copy of the field test training package is available from the first author upon request.

3.5. EHC Field Test Interview Outcomes

The field period for EHC interviewing was from mid-April through the end of June. During this time the FRs completed interviews in 1,627 sample households. With the appropriate discounting of 153 cases deemed ineligible for interview (e.g., those which were found to be vacant, demolished or otherwise unfit for occupancy, converted to business use, etc.), the unit-level response rate for the field test was approximately 91% – see part A of Table 3. Given the nature of the sample, which was dominated by cases which had demonstrated a willingness to be interviewed time after time over several years, we expected (perhaps “hoped for” is more apt) an even higher rate of response. Anecdotal evidence suggests that, while ready cooperation was by no means absent, a more common reaction was annoyance and dismay at the prospect of yet another interview, especially since many households had been assured that the wave 12 SIPP interview was their last.

According to the household rosters obtained in the course of the EHC interview, interviewed households contained a total of 934 children under the age of 15, who were not eligible for an individual EHC interview, and 3,363 interview-eligible “adults” aged 15 and older (these age-eligibility rules match SIPP’s). Individual EHC interviews were completed with 3,318 adults, for a person-level response rate among identified, eligible respondents of almost 99%. These results are summarized in part B of Table 3.

Respondent rules in the EHC field test also mirrored SIPP procedures. The desired form of response was self-response. In cases where self-response was not possible, due to a sample person’s unavailability or unwillingness to respond, FRs were allowed to seek a knowledgeable proxy respondent from among the other adults in the household. About two-thirds of the EHC interviews were administered via self-response; in about one-third of the cases the data were obtained by proxy. These proportions correspond quite closely to what SIPP has achieved historically (see, e.g., U.S. Census Bureau, 1998).

As noted earlier, an after-the-fact clerical procedure matched the people identified in the EHC field test interview to SIPP records to identify those who had, in fact, been the subject of a SIPP interview concerning their 2007 characteristics, and those who had not. Results of this matching operation are summarized in part C of Table 3. Of the 3,318 interviewed adults, a positive match to a SIPP record was achieved for 2,748, or about 83%. The matched cases in the “continuing wave 11” component of the sample comprise the primary analysis sample for the present research. Of the 1,658 cases in this category, 38 were found to have no SIPP interview data for any of the three waves of interest, resulting in a final analysis sample of 1,620 adults (age 15+) interviewed in the EHC field test and in at least one of the SIPP wave 10, 11, or 12 interviews.

3.6. Supplementary Components of the Research Plan

Several additional field test activities provide supplementary information to assist the comparison of the EHC and Q-by-Q interview data and the evaluation of the field test in general. These activities, primarily directed toward a better understanding of the EHC interview process, include a respondent debriefing form, an interviewer debriefing form, interviewer debriefing focus groups, and interview observation reports. We only note these activities here; each is (or

will be) the subject of its own separate report.⁷

4. Research Assumptions and Limitations

We are well aware that the design of this research is far from perfect. Strong assumptions and major limitations include the following:

- Our research sample is small, or at any rate may not be large enough to detect small but important treatment differences for the rare characteristics of most interest to the study (e.g., Food Stamps, SSI, TANF, etc.).
- The sample is drawn from limited areas, and, due to sample aging, nonresponse, and attrition, can no longer be defended as representative of even the population of those limited areas, let alone the general population. The portion of the sample for which we can compare SIPP and EHC reports represents a further curtailment of the sample. These compromises to the sample’s statistical integrity have unknown effects on the generalizability of the field test results.
- The test uses a paper questionnaire, without recourse to dependent procedures, and with a much-reduced set of SIPP-relevant content; the generalizability of the test results to an automated EHC questionnaire with full content is uncertain.
- A third set of generalizability questions concerns respondents. Specifically, the generalizability of the “paper” test results to a sample of new respondents, in a new panel, with no prior SIPP experience, is unknown.
- Another issue, related to the above, concerns the special knowledge of EHC interviewers in some interview situations. We were unsuccessful in designing the test so as to prevent SIPP-experienced FRs from having their own former SIPP cases included in their EHC assignments. Interviewers’ intimate familiarity with the characteristics and circumstances of some of their EHC households may have affected how they administered the EHC instrument in those cases.
- Our test of EHC methods is confounded with a major change in the length of the reference period, from four months in the current SIPP design to a full calendar year with the EHC⁸. If the EHC treatment yields data of lesser (or merely different) quality, the precise cause will be indeterminate.

⁷ At this writing, three of these four reports are complete. Miller (2008) describes observers’ reports; Pascale (2009) summarizes the results of the interviewer debriefing focus groups; and Chan (2009) presents findings from the debriefing forms filled out by respondents in interviewed households.

⁸ This shortcoming is rooted in the fact that the survival of the SIPP program required a major reduction in costs, at a level which could only be achieved through a major reduction in the primary determinant of costs – the number of interviews conducted. A drastic reduction in sample size was deemed untenable, which left increasing the efficiency of each interview as the only viable option. SIPP managers were not interested in exploring a simple expansion of the four-month reference period to twelve months within the traditional SIPP question format, under the assumption that it was not worth the cost and energy to test an option in which the probability of important data quality problems seemed so high.

– The validity of the field test as a test of the EHC approach depends heavily on the assumption that interviewers were well-prepared to implement that approach effectively. Is that assumption tenable? Apart from the contractor’s small pretest of an earlier draft, the field test training package was completely new and untried. In addition, the training was administered to (and by) people who had no prior experience with EHC methods, and for whom many of the key tenets of those methods represented a major departure from familiar rules and practices.

– For some characteristics we will eventually be able to compare survey reports to administrative record data; for the remainder the meaning of any differences between SIPP and EHC reports is uncertain. We assume that particular patterns of differences indicate data quality differences – for example, that an early-in-the-calendar year report of some event in SIPP that is not reported in the EHC interview represents an underreport in the calendar interview, due perhaps to the much greater lag in the EHC between the event date and the report date; or that a mismatch in reported transition dates may be due to seam bias in the SIPP data.

5. Analysis Details

5.1. Characteristics Examined

The focus of this paper is the comparison of the EHC reports of the “continuing wave 11” component of the field test sample with their SIPP reports for the same calendar year 2007 time period. In this initial report, we examine results for eight characteristics of interest to the SIPP program: four “need-based” government transfer programs directed at low income people or families:

Food Stamps,
Supplemental Security Income (SSI),
Temporary Assistance to Needy Families (TANF), and
the Women, Infants, and Children (WIC) nutrition program;

and four other characteristics which are largely independent of respondents’ economic circumstances:

Medicare, the federal government health insurance program primarily for the elderly,
Social Security (OASDI) retirement benefits,
school enrollment, and
employment at a job or business.

In this initial report we focus only on a comparison of yes/no monthly participation (or coverage, or status) reports. Future analyses will address other features of these characteristics, such as month-to-month transitions, and, where appropriate, benefit amounts.

5.2. Data Files

SIPP data for this analysis come from the unedited, internal-use data files that are the first product of the SIPP processing system. These files are one step removed from the raw instrument output, without any of the processing system’s myriad and often complex edits and

imputations which are used to fill data gaps and resolve inconsistencies. The main processing “work” that is carried out at this stage is administrative, rather than substantive – removing duplicate records and identifying and resolving issues that might confound the linking of any individual’s data between subfiles within a wave or across interview waves. We chose to work with these files because they are closest, conceptually, to the files created through the data entry system specially designed to capture the EHC interview results. The guiding principle behind that system was to key the EHC forms “as is,” without trying to fill in missing data or resolve inconsistencies. In fact, our analysis work required that we implement some straightforward edits of the data. For the most part, these were logical edits which replaced missing data with the implied non-missing value, as follows:

(1) Both surveys included “screening” questions which asked respondents whether the characteristic applied to them at all during the reference period (the preceding four months in SIPP; calendar year 2007 in the EHC) – e.g., (for SIPP) “Were you enrolled in school at any time between [month 1] 1st and today?”; or (for the EHC) “Did you receive workers’ compensation at any time during 2007?” When these types of questions elicited a “no” response, detailed questions about month-level details were skipped; under these circumstances, we changed blank monthly indicators to “no.”

(2) The question sequence for some characteristics was designed to skip even the screening question under certain circumstances. In the SIPP, for example, the first spouse to be interviewed is asked about Food Stamps receipt for either the person him/herself or his/her spouse. In the event of a “no” response, the other spouse is not asked the screener question. Another example from SIPP: the screener question for Medicare coverage is skipped entirely for very young (age 15-19) non-disabled adults. Again, under such circumstances, we changed blank monthly indicators to “no.”

(3) In both surveys, reported receipt of some need-based benefits (e.g., Food Stamps, TANF, etc.) elicited follow-up questions to identify other household members also covered by those same benefits. In cases where a reported beneficiary’s own record indicated one or more months of participation, we retained all of the information in that record; in cases where a reported beneficiary’s own record contained no indication of monthly participation, we copied the initial reporter’s monthly data to the beneficiary’s record.

(4) For characteristics associated with benefit amount reports, in both surveys we found occasional instances in which a dollar amount was recorded but the monthly participation indicator was something other than “yes.” In such circumstances we changed the monthly indicator to “yes.”

In general, and equally for both surveys, the editing rules we implemented gave precedence to monthly reports of “yes” (covered, participating, etc.). That is, the presence of a “yes” in a person’s original record resulted in none of the above edits being applied. As a result, our edits only served to increase the number of monthly “yes” reports, and never to reduce them.

5.3. General Analysis Approach

The primary analyses presented in this paper compare the EHC reports of the “continuing wave 11” sample cases with their SIPP reports for the same calendar year 2007 time period. The characteristics we investigate are monthly “participation” reports – yes/no, on/off, covered/not covered, etc. – for each calendar month of 2007. The basic underlying building block of this stage of analysis is a 2x2 contingency table, of the type shown in Figure 2. In an ideal situation, with no missing data, the total N for each month would be 1,620. In our unedited data files, no characteristic is free of some missing data in one or the other (or both) of the surveys. For the January through September period the bulk of missing data results from noninterviews in a particular survey wave. As noted earlier, SIPP’s rotation group structure causes much missing data in the final three months of calendar year 2007.

We employ an exploratory logistic regression modeling approach (using SAS’s GENMOD procedure) as our primary tool for statistical analysis, modeling reported “participation” as a function of interview method (SIPP v. EHC) and time (calendar month); we also add a site factor (IL v. TX) to assess the similarity of the results across the two states in which the field test was carried out. Within the logistic regression approach we use Generalized Estimating Equations (GEE) to account for the high within-person correlation of the responses. (Appendix 4 provides a few more details concerning the modeling approach.) We ignore an individual respondent’s data for any month in which either the SIPP or the EHC report is missing for that month. For the most part, and for most characteristics, missing data for the January-September period is trivial. In October, November, and December, however, the cases available for analysis decline steadily and substantially, as successive SIPP rotation groups complete the 2004 panel, such that only about one-quarter of the SIPP sample remains to report about December – see Figure 1.

5.4. Research Questions

The two primary research questions our analyses seek to address are as follows:

(1) Is there an overall effect of interview treatment – that is, a general tendency for one interview type to obtain more “yes” reports than the other? This we refer to as a main effect for a factor labeled METHOD.

(2) Does the effect of interview type vary across months – i.e., is there a significant METHOD*MONTH interaction? Of particular interest in this regard would be evidence of SIPP>EHC treatment differences early in the year which diminish or disappear later in the year. Although not definitive by any means, such a pattern would be consistent with increased forgetting errors in the EHC, due to the substantially longer recall period for early-in-the-year months – more than a year in the case of the EHC, compared to an average of only two months (and never more than four) in SIPP. For the later months of the year the recall task for the EHC report becomes steadily less demanding, and more comparable to the SIPP situation.

As noted above, we also include a SITE factor in our models to assess whether or not the effects we observe with regard to the two primary research questions are constant across the two field test sites. For the most part, our analyses indicate no impact of SITE, or no important impact. (We view a SITE main effect, indicating an overall difference between IL and TX cases in the reported level of the phenomenon in question, as not important for our purposes.) Thus, in the report of results which follows we mostly ignore the SITE factor. For those instances in which

we do find significant interactions involving SITE, we disaggregate the results and present them separately for IL and TX.

A final note about site differences: Our initial analysis of the TANF results proved to be severely compromised by extremely sparse cells. Specifically, the EHC interview in the IL sample elicited no reports of TANF receipt at all, and there were very, very few in the SIPP interview. As a result of these data problems, in our TANF analysis we ignore the IL component of the sample, and focus solely on the TX cases.

6. Results

6.1. Priming Bias

As noted earlier, the design of the field test justifies legitimate concerns about “priming bias,” due to the fact that EHC respondents from the “continuing wave 11” subsample had already provided information about calendar year 2007 in their SIPP interviews. Is there a risk of drawing the wrong conclusions from the comparisons of SIPP and EHC reports because the latter were “primed” in this manner? Would response to the EHC have been different – more discrepant from the report generated by the standard SIPP approach, perhaps – had those respondents not been interviewed previously about 2007?

We addressed this issue using the wave 8 sample cut component of the field test sample (see section 3.2). These cases reported about calendar year 2007 for the first time in the EHC interview. If there is a bias due to the “priming” effects of the SIPP interviews, the wave 8 sample cut cases should respond differently to the EHC interview than the wave 11 “continuing” cases; if there is no bias then the two groups’ EHC responses should be much the same. In order to make the two samples comparable, we first eliminated from the “continuing” sample all respondents who had moved since their wave 8 interview. (In the wave 8 sample cut component of the field test sample anyone who moved after the wave 8 interview was lost to the EHC interview.) We also applied minor weighting adjustments to the “continuing” sample to force equivalence between the two subsamples on major demographic characteristics⁹.

The results of this analysis strongly support a “no priming bias.” Across all seven¹⁰ characteristics examined, and all months of calendar year 2007, we found the EHC estimates derived from the responses of the wave 8 sample cut respondents to be statistically indistinguishable from those produced by their counterparts in the wave 11 “continuing” sample (data not shown; details available upon request). These results justify greater confidence in the comparisons of SIPP and EHC responses that follow. Those comparisons seem to be untainted by the fact that there was an earlier SIPP interview, and a reasonable approximation of what would have been obtained had “fresh” respondents provided the EHC reports.

6.2. Overall Agreement Between the SIPP and EHC Reports

⁹ Weighting had no impact on the conclusions to be drawn concerning priming bias – weighted and unweighted analyses yielded very similar results.

¹⁰ Due to the almost vanishingly small number of TANF reports, we exclude TANF from the priming bias analysis.

The most prominent feature of the comparison of SIPP and EHC interview reports is that they almost always agree. This statement applies to all monthly comparisons for all of the characteristics examined here. For four of the eight characteristics the two surveys' reports are in agreement over 98% of the time, on average, and for three others the average agreement level is over 96%. Even the poorest-performing characteristic, employment, shows exceptionally high agreement between the SIPP and EHC reports, which agree over 93% of the time, on average, across the twelve months of calendar year 2007.

We carried out a more formal statistical analysis of the level of agreement between the two surveys' reports using Cohen's kappa (Cohen, 1960). The primary advantage of this statistic over the simple percent agreement is that it takes into account the likelihood of agreement by chance. This analysis finds, for every characteristic and every month, kappa values well in excess of the level considered to represent "substantial agreement" (Landis and Koch, 1977); in many instances, in fact, the estimate is well into the "almost perfect agreement" category (results not shown). The point is that, even though the main focus of the analysis is on the extent to which the two survey reports disagree, and the nature of those disagreements, the fact remains that disagreements are quite rare events.

6.3. Examination of SIPP and EHC Report Differences

In this main results section we identify three distinct patterns in the month-by-month comparison of the SIPP and EHC reports for the eight characteristics listed in section 5.1. Each pattern has, we believe, a distinct set of implications concerning the basic question which motivated this research: Can a single EHC interview, with a 12-month, calendar year reference period, obtain data of equivalent quality as three SIPP interviews, each of which covers four months? Thus, we organize the results report around these three observed patterns, rather than by substantive characteristic.

6.3.1. Results Pattern 1 – The SIPP and EHC Estimates are Equivalent All Year

Our logistic regression modeling analysis identifies two characteristics for which the SIPP and EHC reports yield statistically indistinguishable estimates across all twelve months of calendar year 2007. We see this pattern in the results for SSI and WIC¹¹ (IL only) – see Figures 3 and 4¹².

In these figures (and in each of the figures to follow), each estimate in each month's pair of estimates represents the marginal percentage of "yes" (i.e., SSI receipt) reports for that interview

¹¹ Although men can receive WIC benefits, it is quite uncommon for them to do so. The automated SIPP questionnaire employs behind-the-scenes logic to identify certain rare circumstances under which a male might be eligible for WIC benefits, and under those circumstances does ask the WIC screener question of a male respondent. We could not feasibly replicate this logic in a paper-and-pencil interview, and thus the EHC questionnaire instructed interviewers to ask only females about WIC. We therefore exclude males from our analysis of the WIC results.

¹² Appendix 3 contains detailed data tables for these two characteristics, as well as for all of the others described in section 6.3, each of which shows the consistency of the reports derived from the two interviews, including the missing data cases which are excluded from the statistical analyses. Where the statistical analysis identifies SITE as an important factor, we break out the detailed tables by SITE; otherwise we present only the overall results, collapsed across the two states in which the field test was carried out.

treatment among all respondents who provided non-missing data in that month for both interview treatments. Thus, across months, the set of people who provide the data vary somewhat, especially in October, November, and December, as one-quarter, one-half, and three-quarters of the SIPP sample become unavailable for analysis, respectively (see Figure 1). Within each month, however, the data which produce the SIPP and EHC estimates are derived from exactly the same set of people. Below each figure is a summary of the statistical analysis results. In the case of Figures 3 and 4 the formal analysis confirms what is evident to the naked eye – the overall main effect for METHOD is non-significant, and that statement applies to every individual month as well.

We stress that when the two surveys produce identical (or nearly so) estimates, as for these two characteristics, it does not mean that the two survey reports are the same in all cases. Rather, it means that in the monthly comparisons the discrepant cases are equally balanced – i.e., the number of “SIPP yes, EHC no” cases is the same as the number of “SIPP no, EHC yes” cases. In addition, absent an objective yardstick against which to assess the accuracy of people’s reports, we cannot make any statements about the absolute level of data quality in either survey. Whatever the actual level of quality, however, the Pattern 1 results suggest that the quality of the data generated by the two surveys is quite comparable, and that for these two characteristics there is little reason to be concerned about the shift from the current SIPP approach to an EHC-based interview with a 12-month reference period.

6.3.2. Results Pattern 2 – The SIPP and EHC Estimates are Equivalently Different All Year

The second pattern revealed by the statistical analysis is a straightforward (more or less) METHOD main effect – specifically, one in which the SIPP monthly “participation” rate exceeds the EHC rate, and in which the magnitude of the SIPP-EHC difference is essentially constant across all 12 months of 2007. We see this pattern in the results for four characteristics: receipt of Social Security retirement income, Medicare coverage, receipt of WIC benefits (TX only), and receipt of Food Stamps (IL only). In the graphical displays this pattern reveals itself as roughly parallel lines – see Figures 5 through 8.

Keeping in mind all of the appropriate caveats, we find it reasonable to conclude that the Pattern 2 results probably indicate underreporting in the EHC interviews compared to the SIPP reports. While we are comfortable with this conclusion as it applies to the EHC *treatment*, as implemented in the 2008 field test, several factors lead us to suspect that it is at least premature to apply it broad brush to the EHC *method* more generally, or even to the longer reference period with which the EHC method was paired in this test. Perhaps the most compelling argument for this position is the absence of any apparent impact of time – whatever caused the EHC treatment to under-perform was equally problematic early in the year, when the difference between the two interviews in terms of the length of the recall period was at its maximum, and late in the year, when recall length differences between the two survey treatments were minimal. Primarily for this reason we look instead to particular features of the EHC treatment which might have led to a tendency to underreport in-scope events and circumstances in general, rather than through some time-based mechanism. We find several likely suspects, key among which are the following:

- (1) The paper-and-pencil format of the EHC instrument almost necessarily reduced the effectiveness of its screening questions and procedures. For example: the EHC form employed

an abbreviated set of screening questions compared to the SIPP instrument which, for some topics, asks several different forms of a question before accepting a “no” response and moving on to other topics. In addition, of course, EHC interviewers, unlike those using the SIPP instrument, were never faced with *automated* probes or follow-ups, but instead had to find or generate them themselves. The absence of automation in the EHC questionnaire also severely constrained the ability to offer respondents an array of tailored, locally-recognized labels for certain programs. And finally, as noted earlier, the EHC response task was rendered more difficult for respondents due to the absence of any dependent interviewing procedures. In the SIPP interview, many respondents had only to recognize a characteristic of interest, whereas all EHC respondents were forced to rely on much more difficult and error-prone recall memory processes to report on the same characteristics.

(2) The two instruments employed different definitions for some concepts. For example, the EHC questionnaire asked respondents whether they received “Social Security retirement.” The SIPP approach, in contrast, is to ask about receipt of “any Social Security payments,” and later, through a series of follow-up probes, to determine the reason for the payments. Not surprisingly, a more detailed analysis reveals that the vast majority of apparent EHC underreports of Social Security were produced by those who reported in SIPP that they received those payments for a reason other than retirement – disability, most commonly¹³. Thus, although the two surveys produced a clear pattern of discrepant reports about Social Security, in many cases both reports may have been accurate.

(3) The apparent deficit in Medicare reporting in the EHC may be due to the fact that the health insurance series in the EHC questionnaire was particularly complex and troublesome for interviewers, so much so that we suspect that interviewers failed to ask a question about Medicare in some interviews. We attempted to implement an improved design for the health insurance questions (Pascale, 2009a; Pascale, Roemer, and Resnick, 2009), but the resulting script and physical format – in particular, the various paths and follow-ups and even recycling loops for different response scenarios – proved quite challenging. The fact that interviewers found this section of the questionnaire daunting was evident during training, in interview observation reports (Miller, 2008), and in the interviewers’ comments about the EHC experience (Chan [forthcoming]; Pascale, 2009b).

6.3.3. Results Pattern 3 – SIPP Estimates Exceed the EHC, but Only Early in the Year

The distinguishing feature of the third basic pattern evident in the SIPP-EHC comparison is the significant impact of time, and in fact a particular form of that impact – a marked deficit in EHC reports relative to SIPP early in the year, followed by the elimination of that deficit late in the year. Four characteristics display this pattern – Food Stamps (TX only), TANF (TX only), employment, and school enrollment; see Figures 9 through 12. The precise definitions of “early in the year” and “late in the year” show some variation across the four characteristics, and there is variation as well in the “post-deficit” patterns of SIPP-EHC differences. But all four characteristics show clear evidence of potential underreporting in the EHC, relative to SIPP, in

¹³ For example, according to their receipt detail reports in SIPP, 22 of the 26 respondents with “SIPP=yes, EHC=no” discrepancies in January were receiving Social Security for reasons other than retirement. Similar patterns are evident for all months (data not shown).

the January through May period (and beyond, in some cases), and in all four cases that problem is gone by October (well before October, in most cases).

Pattern 3 suggests that concerns about data quality in a re-engineered SIPP interview with a 12-month reference period are not baseless, and that the demonstrated strengths of the EHC method may not be sufficient to overcome the additional recall difficulties for the longest-ago months of the reference period. Of additional concern is the fact that two of the characteristics which display this pattern are the very same need-based programs for which prior research has found EHC methods to be no better than a standard Q-by-Q questionnaire (Belli, Shay, and Stafford, 2001).

7. Discussion, Conclusions, and Next Steps

Overall, we view the 2008 field test experience as a very successful “proof of concept” for the use of EHC methods in an expanded reference period in the interview for the re-engineered SIPP, and certainly sufficiently successful to proceed with a larger test of an automated EHC instrument, as is currently planned for early 2010. Despite the many limitations of the field test design, and despite what would seem to be a large disadvantage for the EHC treatment in terms of the span of time over which key facts and dates need to be retained in memory, we find, in general, very high levels of agreement between the EHC and SIPP reports. But the specific results are only part of the story – understanding and quantifying questionnaire design effects on subject-matter data quality may have been the most important goal of the 2008 exercise, but it was not the only important goal. Other considerations, to which we devote little or no attention in this paper, also lead us to label the test a success. The field test also served as an opportunity to gain vital operational insights – for example, about the importance of the timing of the field period for an annual survey (an issue that SIPP has never had to consider before), and how to train interviewers to carry out EHC procedures, and how to implement key features of those procedures, such as capturing and using landmark events. In these respects and many others, the 2008 field test both demonstrated that the re-engineered SIPP plan could be implemented successfully, and at the same time pointed the way to how that plan could be implemented better.

With regard to the specific results of our analyses presented here, the yield can only be described as a somewhat mixed bag. It seems quite clear that part of the variability in the results has to do with whether or not the characteristic in question is a state-administered program. The analyses for all three of those programs considered in this paper – Food Stamps, TANF, and WIC – yielded significant interactions with the SITE (state) factor; this was the case for none of the other characteristics. This suggests that state-based differences in program names or program administration details may have effects on recipients’ abilities to report them accurately in a survey.

But there is more to the “mixed bag” of results than variation within state-based programs – we see a wide range of outcomes across all of the characteristics, with a similarly wide range of implications. For some characteristics (SSI; WIC (IL)) the two survey treatments produced response profiles which are nearly indistinguishable, and for those characteristics our conclusion is that there is very little reason to be concerned about the impact on data quality of the re-engineered SIPP’s proposed design. Several others (Social Security; Medicare; WIC (TX); Food Stamps (IL)) display a pattern of constant treatment differences across the calendar year, which

probably signals some data quality issues for the experimental treatment in the field test, but not necessarily for an annual EHC interview with a 12-month reference period, as the re-engineering effort envisions. We suspect that refinements to the EHC instrument currently being developed for the 2010 test – and especially its automation – will go a long way toward solving these problems. This will be a key area of interest in analyzing the results of the next test.

The final pattern of results – the pattern that finds reduced reporting in the EHC treatment compared to SIPP, but only in the early months of the year – is the most troubling. We see this pattern in the results for Food Stamps (TX), TANF (TX), paid employment, and school enrollment. This pattern is troubling because it is consistent with not-unreasonable concerns that a survey with a 12-month reference period, compared to a 4-month reference period, might suffer a data quality “hit” for the parts of the reference period most distant from the time of the interview. Clearly, evidence regarding this same sort of problem will also need to be a major focus of the analysis of the results of the next field test, in 2010.

However, even before that next test, there may be more that we can do with the data from the 2008 test to better understand this phenomenon. First, for example, are there some features of these particular characteristics which render them more vulnerable to recall decay? Might those same features, if better understood, yield insights into more effective questions? And second, are the apparent problems real? What do administrative records suggest about data quality differences? And how much of the EHC’s apparent trouble is due to the later-than-optimal timing of the 2008 test? Would EHC interviews conducted immediately after the start of the year have fared better in the early months of the reference period? We may be able to use the existing data to shed light on some of these issues.

The results presented here only begin to scratch the surface of the information potential of the 2008 SIPP-EHC field test. There are additional characteristics to be analyzed in the same manner as those described in this paper: reports of employment at a specific job, spells of looking for work, unpaid labor, receipt of unemployment benefits, receipt of workers’ compensation, non-Medicare health insurance coverage (both specific forms of coverage, and insured/uninsured spells), and asset ownership. The timing of transitions is an important issue for SIPP; we have yet to compare the two surveys with regard to reports of change from one month to the next. In addition to the mere fact of participation in the program, for some programs we also have reports from both surveys of the benefit amount received, and we have not yet looked at those data. We are still in the process of obtaining administrative record data for a subset of characteristics measured in both surveys, so that we can eventually apply an objective standard of “truth” to the comparison of the reports. In short, much more work remains beyond the basic analyses reported in this paper.

References

- Belli, R. (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., Stafford, F., and Alwin, D. (eds.) (2008), *Calendar and Time Diary Methods in Life Course Research*. Thousand Oaks, CA: Sage Publications.
- Belli, R., Shay, W., and Stafford, F. (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
- Caspi, A., Moffitt, T., Thornton, A., Freedman, D., Amell, J., Harrington, H., Smeijers, J., and Silva, P. (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.
- Chan, A. (forthcoming), "The 2008 SIPP Event History Calendar (EHC) Field Test: Analysis of FR Debriefing Questionnaires" [tentative title], Washington, DC: US Census Bureau, *Statistical Research Division Research Report Series*.
- Chan, A. (2009), "The 2008 SIPP Event History Calendar (EHC) Field Test: Respondents' Reactions to the Interview," Washington, DC: US Census Bureau, *Statistical Research Division Research Report Series (Survey Methodology #2009-03)*, issued April 20, 2009; available online at <<http://www.census.gov/srd/papers/pdf/rsm2009-03.pdf>>.
- Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, 20 (1), 37-46.
- Ensel, W., Peek, K., Lin, N., and Lai, G. (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," paper prepared for presentation at the Federal Committee on Statistical Methodology Statistical Policy Seminar, November 2007.
- Fields, J. and Moore, J. (2007), "Description of Plans for a SIPP Calendar Validation Study: Study Design and Analysis," paper presented at the U.S. Census Bureau / Panel Study of Income Dynamics Event History Calendar Research Conference paper, December 2007.

- Flanagan, P. (2006), "SIPP 2004 Panel Wave 9 Sample Reduction (SAMP-1)," unpublished U.S. Census Bureau memorandum to J. Mackey Lewis, July 21, 2006.
- Freedman, D., Thornton, A., Camburn, D., Alwin, D., and Young-DeMarco, L. (1988), "The Life History Calendar: A Technique for Collecting Retrospective Data," *Sociological Methodology*, 18, 37-68.
- Kominski, R. (1990), "The SIPP Event History Calendar: Aiding Respondents in the Dating of Longitudinal Events," Washington D.C.: American Statistical Association, Proceedings of the Section on Survey Research Methods, 553-558.
- Landis, J., and Koch G. (1977), "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, 33 (1), 159-174.
- Miller, D. (2008), "Summary of Observers' Reports from the Re-Engineered Survey of Income and Program Participation Event History Calendar Field Test," Washington, DC: US Census Bureau, Study Series (Survey Methodology) #2008-16, issued November 21, 2008.
- Moore, J. (2008), "Seam Bias in the 2004 SIPP Panel: Much Improved, but Much Bias Still Remains," Washington, DC: US Census Bureau, Research Report Series (Survey Methodology) #2008-3, issued February 25, 2008.
- Moore, J. and Marquis, K. (1989), "Using Administrative Record Data to Evaluate the Quality of Survey Estimates," *Survey Methodology* 15: 129-143.
- Pascale, J. (2009a), "Assessing Measurement Error in Health Insurance Reporting: A Qualitative Study of the Current Population Study," *Inquiry* 45 (5): 422-437.
- Pascale, J. (2009b), "Event History Calendar Field Test Field Representative Focus Group Report," Washington, DC: US Census Bureau, *Statistical Research Division Study Series (Survey Methodology #2009-02)*, issued February 24, 2009; available online at <<http://www.census.gov/srd/papers/pdf/ssm2009-02.pdf>>.
- Pascale, J., Roemer, M., and Resnick, D. (2009), "Medicaid Underreporting in the CPS: Results from a Record Check Study," *Public Opinion Quarterly*, (in press).
- U.S. Census Bureau (1998), "SIPP Quality Profile 1998," SIPP Working Paper Number 230, 3rd Edition.
- U.S. Census Bureau (2001), Survey of Income and Program Participation Users' Guide, 3rd edition, Washington, DC: U.S. Census Bureau.
- Yoshihama, M., Gillespie, B., Hammok, A., Belli, R., and Tolman, R. (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.

Appendix 1 – FIGURES 1 THROUGH 12

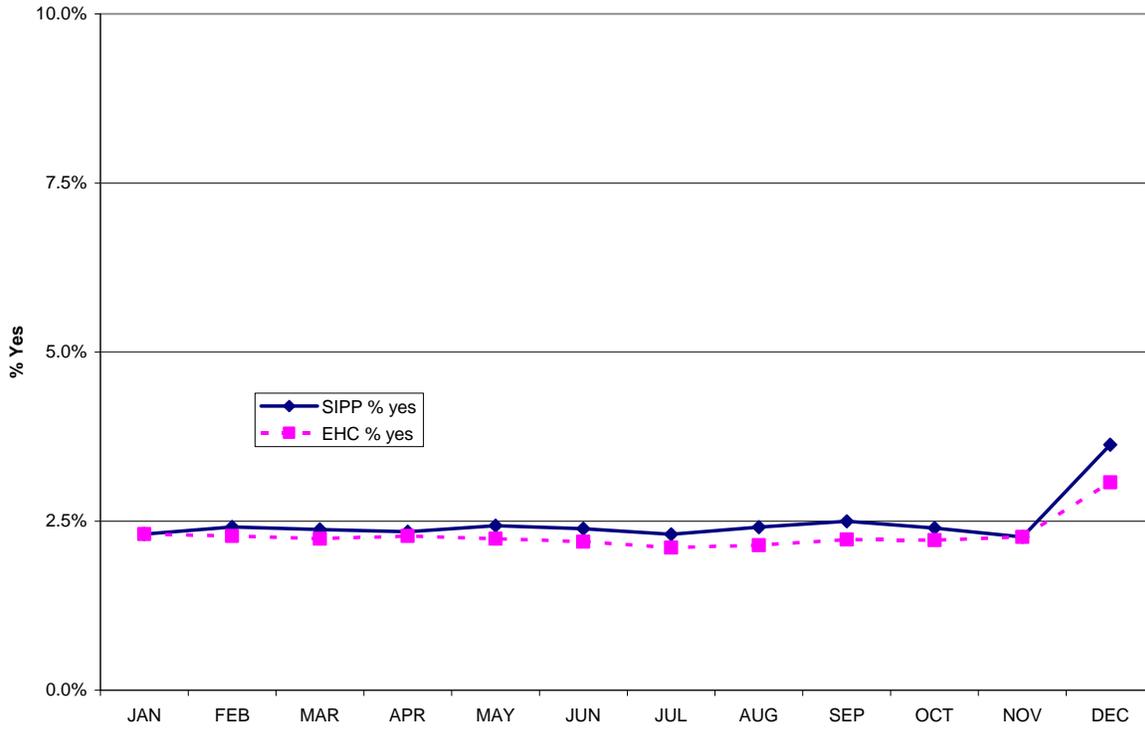
Figure 1: SIPP 2004 Panel Reference Period Months In Calendar Year 2007 By Rotation Group									
CALENDAR MONTH		ROTATION GROUP							
		1		2		3		4	
		Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month
2006	October	W10							
	November								
	December		W10						
2007	JANUARY	W11			W10				
	FEBRUARY		W10					W10	
	MARCH				W10				
	APRIL	W12		W11			W10		
	MAY								W10
	JUNE		W11			W11			
	JULY	W12							
	AUGUST				W11				
	SEPTEMBER						W11		
	OCTOBER	W12	W12						
	NOVEMBER				W12				
	DECEMBER						W12		
2008	January								W12

Note: Within each rotation group, shaded areas show the months of 2007 covered by the wave 10, 11, and 12 SIPP interviews, and cross-hatched areas show months of 2007 for which no data were collected.

Figure 2: SIPP and EHC Survey Reports of [Characteristic X] in [Month] for EHC Interview Cases Matched to SIPP			
		SIPP	
		No	yes
EHC	no	a	b
	yes	c	d

N

Figure 3:
SSI Receipt in 2007 as Reported in SIPP and in the EHC
by Month, Collapsed Across Site
 (see the Appendix 3 data table for details)

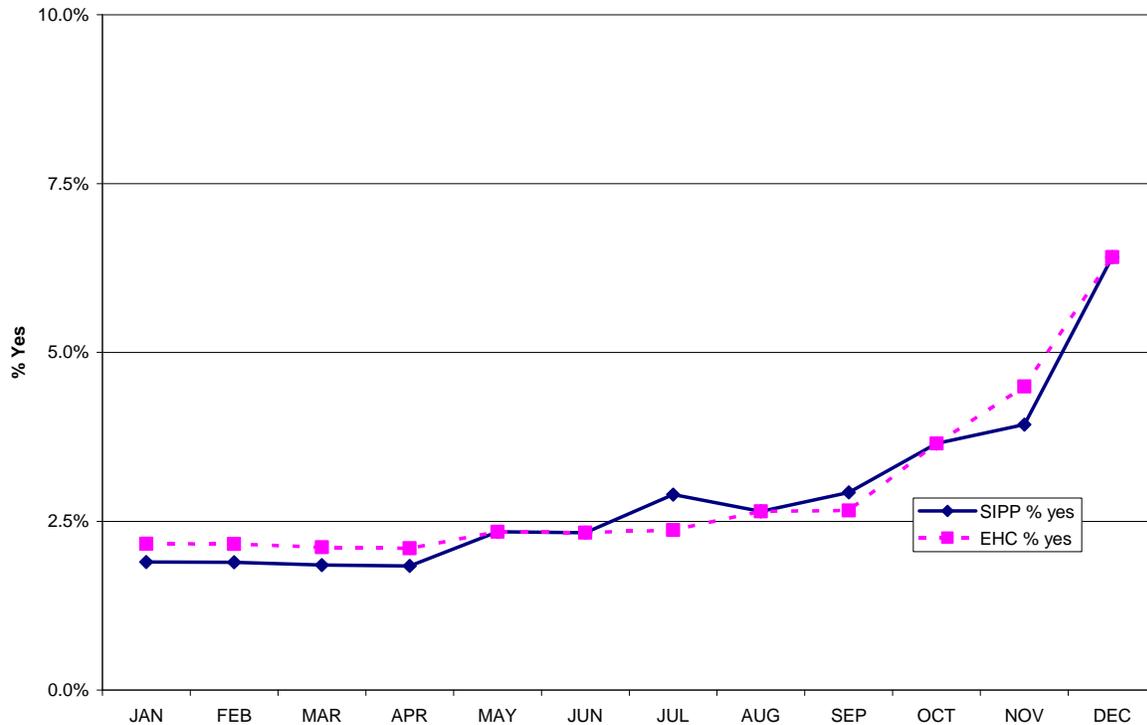


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 1 [see text]:

- no significant METHOD main effect ($z = -0.56, p = .5751$)
- no month's METHOD difference differs significantly from the reference month (JAN)
- no individual month's METHOD difference differs significantly from zero

Figure 4:
WIC Receipt in 2007 as Reported in SIPP and in the EHC
by Month, in ILLINOIS
 (see the Appendix 3 data table for details)

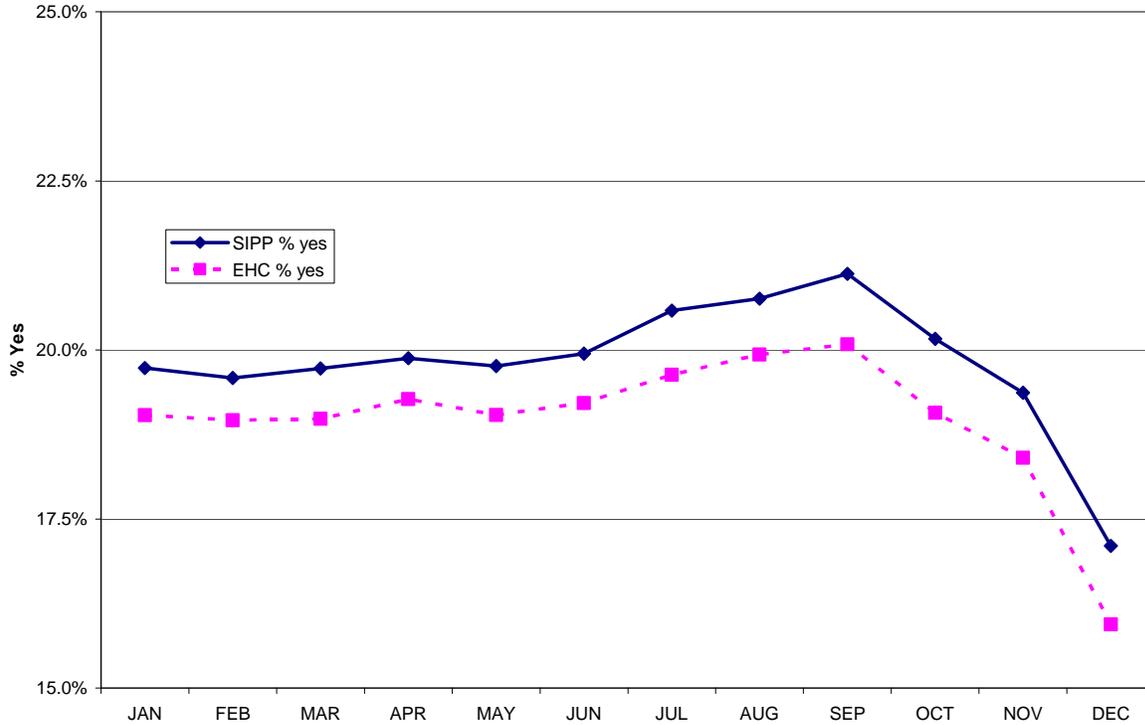


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 1:

- no significant METHOD main effect ($z = 0.23, p = .8202$)
- only the JUL METHOD difference differs significantly from the reference month (JAN) ($z = 1.79, p = .0727$); no other month's METHOD difference differs from the reference month
- no individual month's METHOD difference differs significantly from zero

Figure 5:
Receipt of Social Security Retirement in 2007 as Reported in SIPP and in the EHC
by Month, Collapsed Across Sites
 (see the Appendix 3 data table for details)

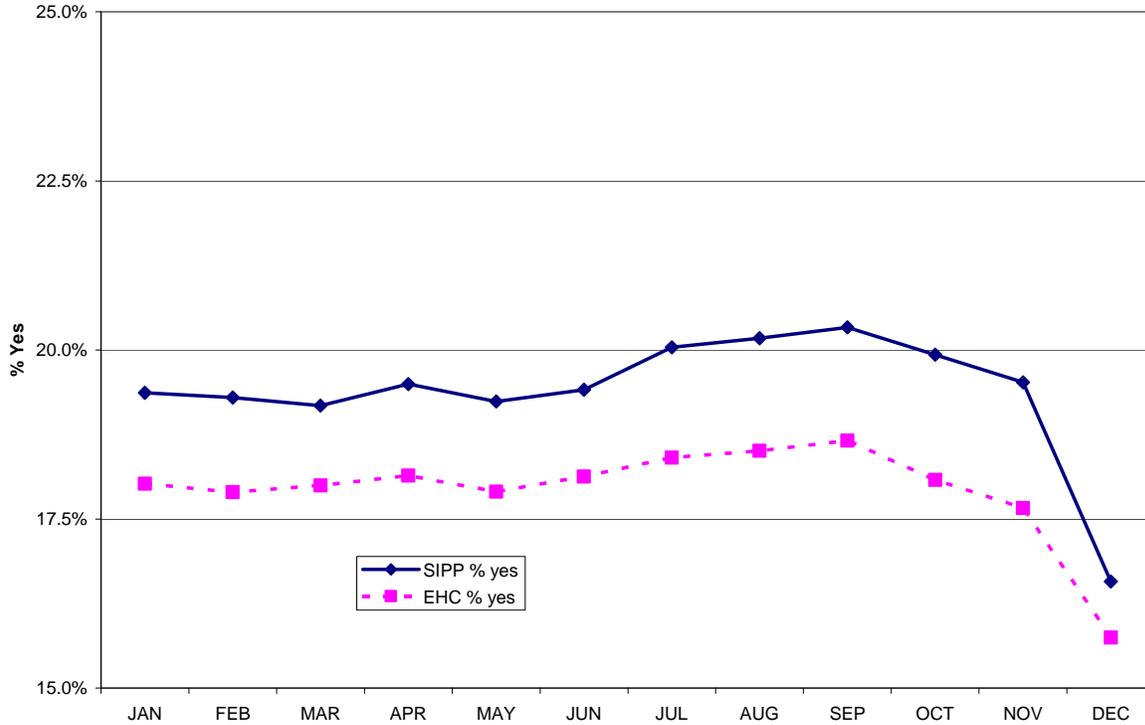


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 2:

- significant METHOD main effect ($z = -1.99, p = .0464$)
- no month's METHOD difference differs from the reference month (JAN)
- most individual months' METHOD differences differ significantly from zero ($p < .10$); all are at least borderline

Figure 6:
Medicare Coverage in 2007 as Reported in SIPP and in the EHC
by Month, Collapsed Across Site
 (see the Appendix 3 data table for details)

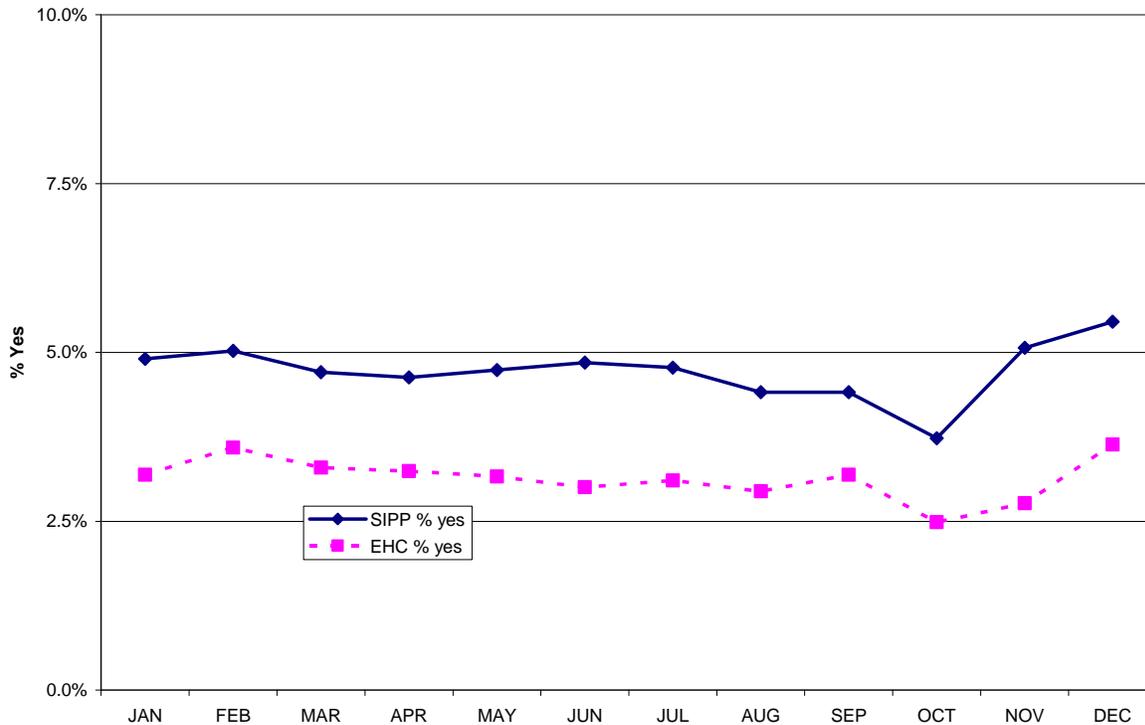


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 2:

- significant METHOD main effect ($z = -3.14, p = .0017$)
- no month's METHOD difference differs from the reference month (JAN)
- except for DEC, all individual months' METHOD differences differ significantly from zero ($p < .05$)

Figure 7:
WIC Receipt in 2007 as Reported in SIPP and in the EHC
by Month, in TEXAS
 (see the Appendix 3 data table for details)

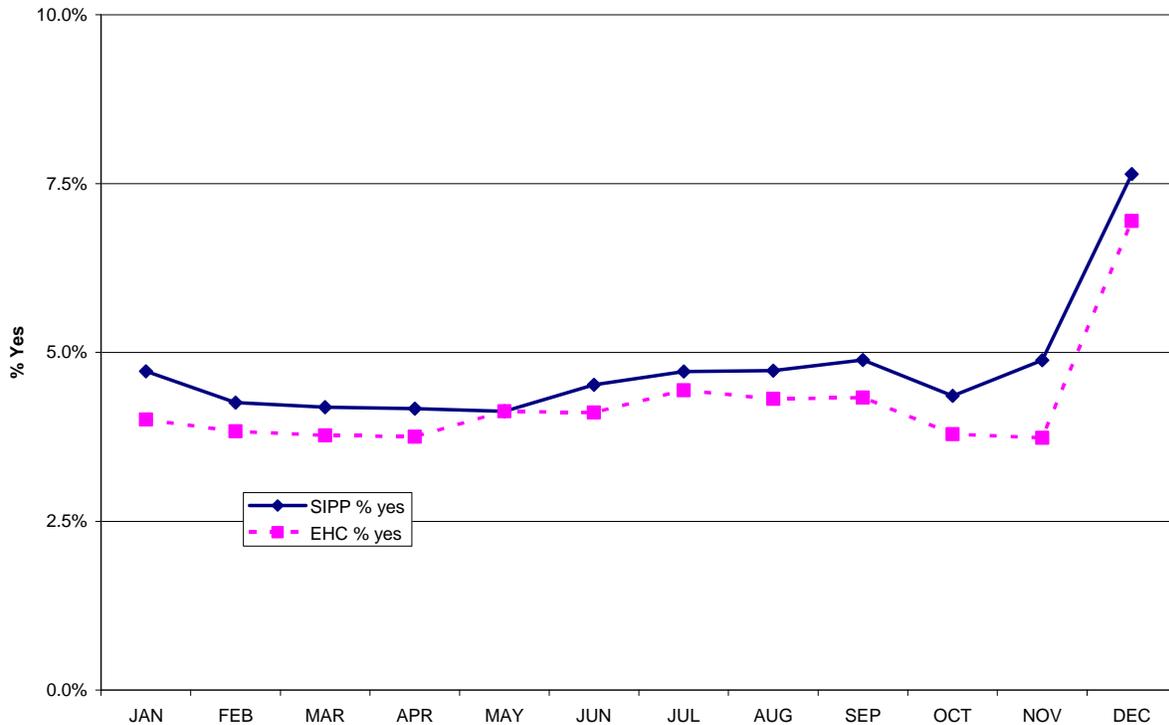


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 2:

- significant METHOD main effect ($z = -2.41, p = .0158$)
- no month's METHOD difference differs from the reference month (JAN)
- most individual months' METHOD differences differ significantly from zero ($p < .10$); all are at least borderline

Figure 8
Food Stamps Receipt in 2007 as Reported in SIPP and in the EHC
by Month, in ILLINOIS
 (see the Appendix 3 data table for details)

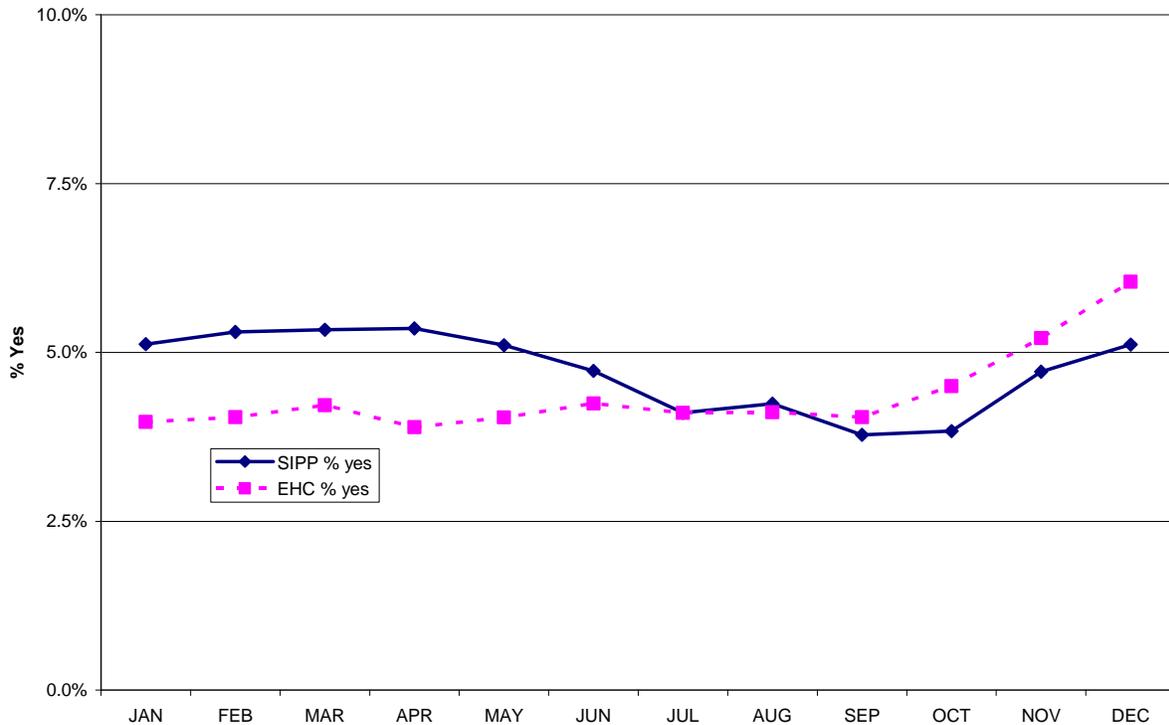


Summary of Statistical Analysis (exploratory logistic regression)

Pattern 2:

- significant METHOD main effect ($z = -1.91, p = .0559$)
- no month's METHOD difference (except MAY) differs from the reference month (JAN)
- most individual months' METHOD differences differ significantly from zero ($p < .10$); all (except MAY) are at least borderline

Figure 9
Food Stamps Receipt in 2007 as Reported in SIPP and in the EHC
by Month, in TEXAS
 (see the Appendix 3 data table for details)



Summary of Statistical Analysis (exploratory logistic regression)

Pattern 3:

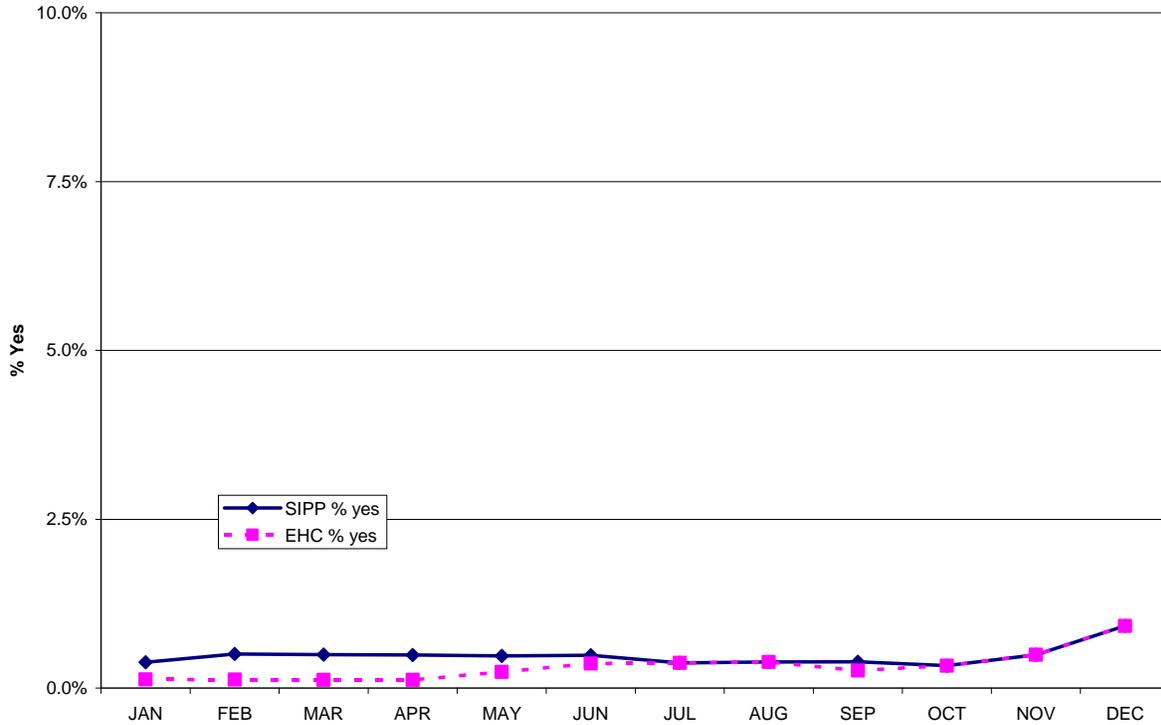
- no significant METHOD main effect ($z = -1.08$, $p = .2808$)

- significant METHOD differences by month:

JAN through JUN: JAN difference differs significantly from zero ($p = .0515$); no other month's difference differs from the reference month (JAN)

JUL through DEC: all months' differences differ from the reference month

Figure 10
TANF Receipt in 2007 as Reported in SIPP and in the EHC
by Month, in TEXAS
 (see the Appendix 3 data table for details)



Summary of Statistical Analysis (exploratory logistic regression)

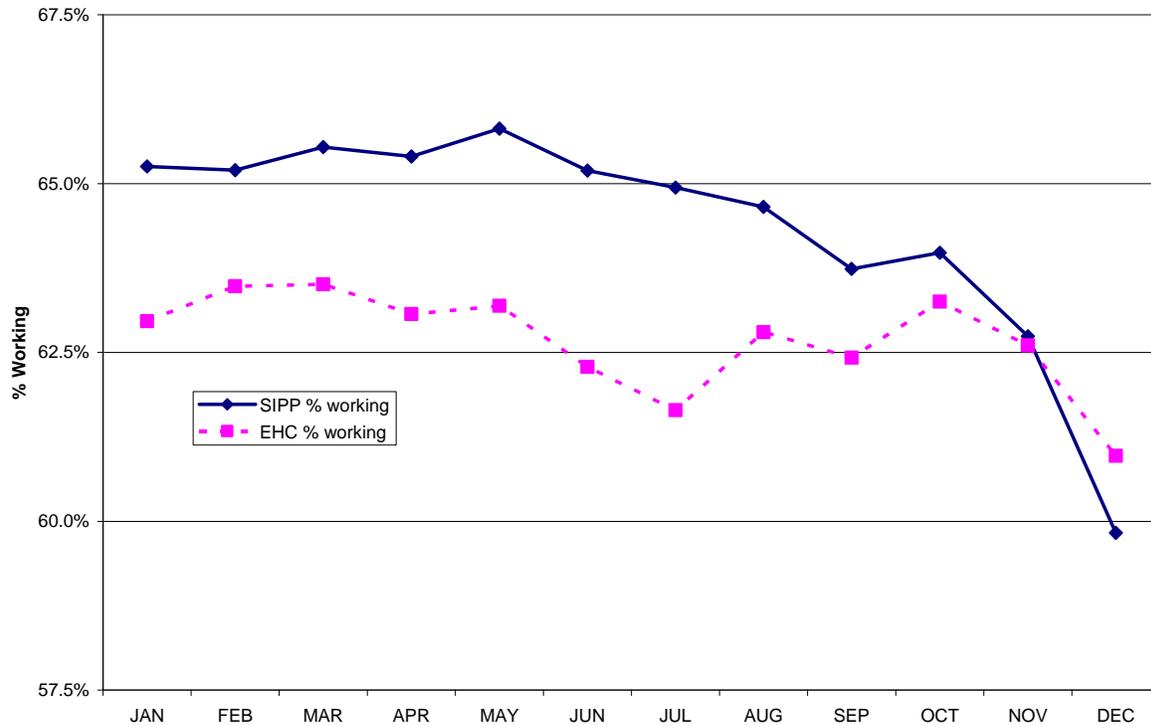
Pattern 3:

- no significant METHOD main effect ($z = -1.31, p = .1911$)
- significant METHOD differences by month:

JAN through MAY: JAN difference differs significantly from zero ($p < .10$); no other month's difference differs from the reference month (JAN)

JUN through DEC: all months' differences differ from the reference month; no month's difference is significant

Figure 11
Any Work for Pay in 2007 as Reported in SIPP and in the EHC
by Month, Collapsed Across Site
 (see the Appendix 3 data table for details)



Summary of Statistical Analysis (exploratory logistic regression)

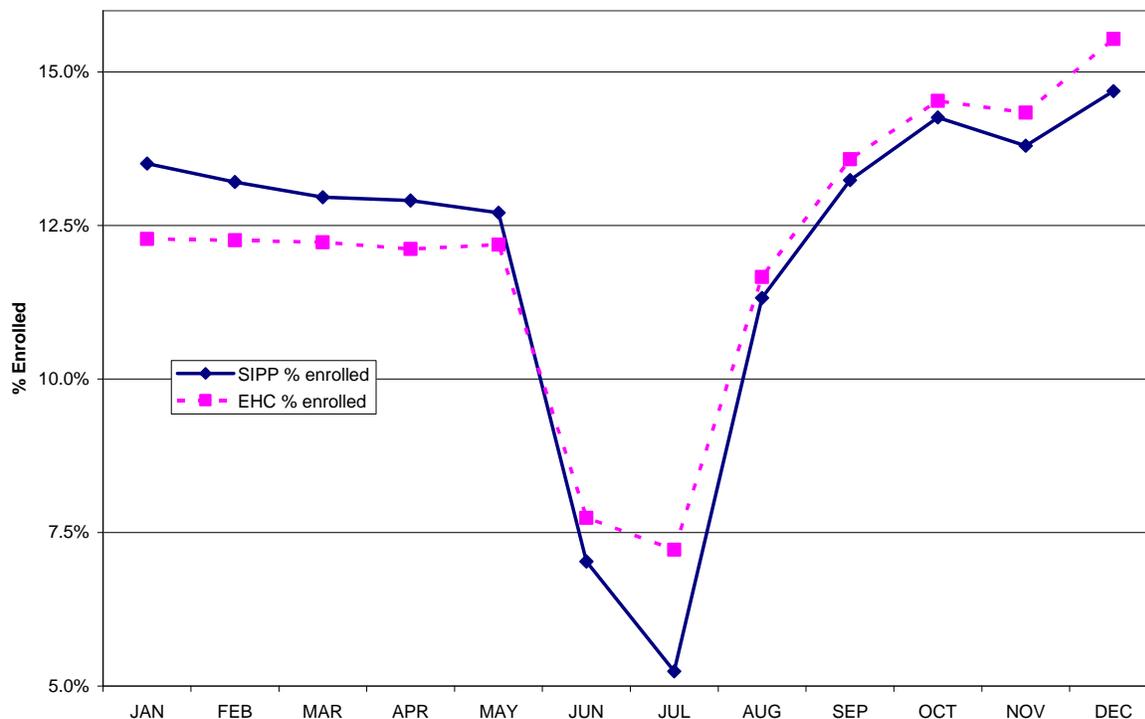
Pattern 3:

- significant METHOD main effect ($z = -3.93, p < .0001$)
- significant METHOD differences by month:

JAN through SEP: JAN difference differs significantly from zero ($p = .0013$), as do the differences for all other months; no month's difference differs from the reference month (JAN)

OCT through DEC: all months' differences differ significantly from the reference month; no month's difference differs significantly from zero

Figure 12
School Enrollment in 2007 as Reported in SIPP and in the EHC
by Month, Collapsed Across Site
 (see the Appendix 3 data table for details)



Summary of Statistical Analysis (exploratory logistic regression)

Pattern 3:

- no significant METHOD main effect ($z = -0.04$, $p = .9690$)

- significant METHOD differences by month:

JAN through MAY: JAN difference differs significantly from zero ($p = .0049$), as do the differences for each of the other months (except MAY); no month's difference differs from the reference month (JAN); significant SIPP>EHC METHOD difference for JAN through MAY combined ($z = -2.18$, $p = .0294$)

JUN through DEC: all months' differences differ from the reference month (JAN); significant EHC>SIPP METHOD difference for JUL ($z = 3.34$, $p = .0008$); significant EHC>SIPP METHOD difference for JUN through DEC combined ($z = 1.88$, $p = .0599$)

Appendix 2: TEXT TABLES 1 THROUGH 3

Table 1: EHC Field Test Sample Size and Composition, by State			
ILLINOIS n = 914		TEXAS (4 METRO AREAS) n = 1,031	
Continuing W11	W8 Sample Cut	Continuing W11	W8 Sample Cut
487	427	609	422

Table 2: Numbers and Characteristics of EHC Field Test Interviewers, by State			
	Illinois (Chicago RO)	Texas (Dallas RO)	TOTAL n (%)
Experienced – Census/ACS only	7	32	39 (36%)
Experienced – other demographic surveys (non-SIPP)	10	13	23 (21%)
SIPP experienced (2+ years)	13	13	26 (24%)
SIPP experienced (< 2 years)	0	5	5 (5%)
New hires for the EHC field test	14	0	14 (13%)
TOTAL	44	63	107 (100%)

Table 3: EHC Field Test Household- and Person-Level Interview Outcomes and Matching Results, by State and Sample Component				
	ILLINOIS		TEXAS	
	Continuing W11	W8 Sample Cut	Continuing W11	W8 Sample Cut
<u>A. INTERVIEW ATTEMPT OUTCOMES</u>				
TOTAL N (addresses)	487	427	609	422
Ineligible	30	39	38	46
nonresponse ("Type A")	40	41	53	31
EHC Interview Cases	417	347	518	345
overall household-level response rate: (1,627 completed interviews / 1,792 eligible cases) * 100 = 90.8%				
<u>B. PEOPLE IN INTERVIEWED HOUSEHOLDS</u>				
TOTAL PEOPLE	1,113	905	1,376	903
Children (<15)	236	182	306	210
Adults (15+)	877	723	1,070	693
Interviewed Adults	866	707	1,056	689
overall person-level response rate: (3,318 completed interviews / 3,363 listed adults * 100 = 98.7%				
<u>C. MATCH-TO-SIPP RESULTS</u>				
All Adults:				
1 - definite MATCH to a SIPP person	765	589	878	493
2 - probable match	10	4	22	12
3 - possible match	-	-	8	4
4 - definite NON-MATCH	101	130	162	184
Interviewed Adults:				
1 - definite MATCH to a SIPP person	758	587	869	491
2 - probable match	9	1	22	11
3 - possible match	-	-	4	4
4 - definite NON-MATCH	99	119	161	183
match-to-SIPP rate among interviewed adults: (match codes 1 and 2) / 3,318 * 100 = 2,748 / 3,318 * 100 = 82.8%				

Appendix 3: DETAILED DATA TABLES

Data table for Figure 3:

SSI Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, Collapsed Across Site										
Note: N=1,620. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP					SIPP			
		?	N	Y			?	N	Y	
E V E N T H I S T O R Y C A L E N D A R	?	1	11	0	J	?	1	11	0	J
	N	133	1429	10	A	?	87	1474	12	U
	Y	2	10	24	N	?	3	9	23	L
	?	1	11	0	F	?	1	11	0	A
	N	115	1446	11	E	?	113	1449	11	U
	Y	2	9	25	B	?	3	7	25	G
	?	1	11	0	M	?	0	12	0	S
	N	90	1471	11	A	?	124	1436	12	E
	Y	2	9	25	R	?	3	8	25	P
	?	1	11	0	A	?	1	11	0	O
	N	72	1489	11	P	?	471	1093	8	C
	Y	1	10	25	R	?	11	6	19	T
?	1	11	0	M	?	5	7	0	N	
N	45	1515	12	A	?	838	727	6	O	
Y	1	9	26	Y	?	20	6	11	V	
?	1	11	0	J	?	8	4	0	D	
N	58	1503	12	U	?	1224	342	5	E	
Y	1	9	25	N	?	26	3	8	C	

Data table for Figure 4:

WIC Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, in ILLINOIS										
Note: Females only; N=404. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP			SIPP					
		?	N	Y						
E V E N T H I S T O R Y C A L E N D A R	?	0	14	0	J	?	0	14	0	J
	N	20	361	0	A	10	369	2	U	
	Y	1	1	7	N	0	0	9	L	
	?	0	14	0	F	0	14	0	A	
	N	19	362	0	E	12	367	1	U	
	Y	1	1	7	B	0	1	9	G	
	?	0	14	0	M	0	14	0	S	
	N	11	370	0	A	14	365	1	E	
	Y	1	1	7	R	0	0	10	P	
	?	0	14	0	A	5	9	0	O	
	N	8	373	0	P	115	263	1	C	
	Y	1	1	7	R	1	1	9	T	
	?	0	14	0	M	7	7	0	N	
	N	6	374	1	A	208	169	1	O	
	Y	0	1	8	Y	4	2	6	V	
	?	0	14	0	J	11	3	0	D	
	N	4	376	1	U	305	72	1	E	
	Y	0	1	8	N	7	1	4	C	

Data table for Figure 5:

Social Security Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, Collapsed Across Site											
Note: N=1,620. Missing data (shaded cells) are ignored in the statistical analysis.											
		SIPP					SIPP				
		?	N	Y			?	N	Y		
E V E N T	?	2	4	5	J		1	1	4	J	
	N	155	1131	26	A		125	1160	27	U	
	Y	25	16	256	N		12	13	277	L	
	?	2	4	5			0	1	3		
	N	141	1146	25	F		151	1136	25	A	
	Y	23	16	258	B		15	13	276	U	
	?	1	4	5			0	1	3		
	N	121	1166	25	M		162	1124	26	S	
	Y	19	14	265	A		15	11	278	E	
	?	1	2	5			1	0	1		
	N	110	1177	25	P		425	866	21	O	
	Y	13	16	271	R		96	9	200	C	
H I S T O R Y	?	1	2	4			3	0	0		
	N	83	1202	27	M		718	581	13	N	
	Y	12	16	273	A		171	6	128	O	
	?	1	2	4			3	0	0		
	N	97	1188	27	Y		1022	284	6	D	
	Y	12	16	273			250	2	53	E	
	?	1	2	4	J						
	N	97	1188	27	U						
	Y	12	16	273	N						
	C A L E N D A R	?	1	2	4						
		N	97	1188	27						
		Y	12	16	273						
?		1	2	4							
N		97	1188	27							
Y		12	16	273							
?		1	2	4							
N		97	1188	27							
Y		12	16	273							
?		1	2	4							
N		97	1188	27							
Y		12	16	273							

Data table for Figure 6:

Medicare Enrollment in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, Collapsed Across Site											
Note: N=1,620. Missing data (shaded cells) are ignored in the statistical analysis.											
		SIPP						SIPP			
		?	N	Y		?	N	Y			
E V E N T H I S T O R Y C A L E N D A R	?	1	3	1	J A N	0	4	1	J U L		
	N	113	1184	35		80	1210	40			
	Y	15	15	253		3	15	267			
	?	1	3	1	F E B	0	4	1	A U G		
	N	98	1199	35		105	1185	39			
	Y	14	14	255		8	14	264			
	?	1	3	1	M A R	1	3	1	S E P		
	N	78	1218	35		117	1173	39			
	Y	9	17	258		8	14	264			
	?	1	3	1	A P R	1	3	1	O C T		
	N	63	1230	38		399	900	29			
	Y	3	17	264		82	8	197			
?	0	4	1	M A Y	1	3	1	N O V			
N	38	1254	39		708	603	17				
Y	2	18	264		154	3	130				
?	0	4	1	J U N	5	0	0	D E C			
N	52	1240	38		1024	301	4				
Y	2	18	265		229	1	56				

Data table for Figure 7:

WIC Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, in TEXAS										
Note: Females only; N=471. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP					SIPP			
		?	N	Y			?	N	Y	
E V E N T H I S T O R Y C A L E N D A R	?	2	12	1	J A N	?	1	13	1	J U L
	N	47	386	9		37	397	9		
	Y	1	2	11		0	2	11		
	?	2	12	1	F E B	2	13	0	A U G	
	N	38	395	8		47	389	7		
	Y	0	2	13		1	1	11		
	?	1	13	1	M A R	2	13	0	S E P	
	N	31	403	8		47	388	7		
	Y	0	2	12		1	2	11		
	?	0	14	1	A P R	6	9	0	O C T	
	N	24	410	8		128	308	6		
	Y	0	2	12		6	2	6		
	?	0	14	1	M A Y	8	7	0	N O V	
	N	13	420	9		231	205	6		
	Y	0	2	12		8	1	5		
	?	0	14	1	J U N	12	3	0	D E C	
	N	23	410	10		336	103	3		
	Y	0	2	11		10	1	3		

Data table for Figure 8:

Food Stamps Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, in ILLINOIS										
Note: N=750. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP					SIPP			
		?	N	Y			?	N	Y	
E V E N T H I S T O R Y C A L E N D A R	?	0	4	0	J	?	0	3	1	
	N	45	665	6	A	?	24	684	5	
	Y	2	1	27	N	?	1	3	29	
	?	04	9	0	F	?	0	4	0	
	N	37	672	6	E	?	26	683	5	
	Y	4	3	24	B	?	1	2	29	
	?	0	4	0	M	?	0	4	0	
	N	26	683	6	A	?	29	680	5	
	Y	4	3	24	R	?	1	1	30	
	?	0	4	0	A	?	0	4	0	
	N	22	687	6	P	?	206	504	4	
	Y	4	3	24	R	?	12	1	19	
	?	0	4	0	M	?	2	2	0	
	N	17	693	4	A	?	379	331	4	
	Y	2	4	26	Y	?	19	0	13	
	?	0	3	1	J	?	3	1	0	
	N	14	695	5	U	?	578	133	1	
	Y	2	2	28	N	?	24	0	10	
										J U L
										A U G
										S E P
										O C T
										N O V
										D E C

Data table for Figure 9:

Food Stamps Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, in TEXAS							
Note: N=870. Missing data (shaded cells) are ignored in the statistical analysis.							
		SIPP					
		?	N	Y			
E V E N T	?	0	5	0	J		
	N	79	733	17	A		
	Y	5	8	23	N		
	?	0	5	0	F		
	N	68	742	18	E		
	Y	5	8	24	B		
	?	0	5	0	M		
	N	56	755	17	A		
	Y	3	8	26	R		
	H I S T O R Y	?	0	5	0	A	
N		40	771	19	P		
Y		3	7	25	R		
?		0	5	0	M		
N		23	792	16	A		
Y		0	7	27	Y		
?		0	5	0	J		
N		40	778	12	U		
Y		0	8	27	N		
C A L E N D A R		?	0	5	0	J	
	N	40	778	12	U		
	Y	0	8	27	N		
	?	0	5	0	J		
	N	40	778	12	U		
	Y	0	8	27	N		
	?	0	5	0	J		
	N	40	778	12	U		
	Y	0	8	27	N		
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y	0	8	27	N			
?	0	5	0	J			
N	40	778	12	U			
Y							

Data table for Figure 10:

TANF Receipt in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, in TEXAS										
Note: N=870. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP						SIPP		
		?	N	Y		?	N	Y		
E V E N T H I S T O R Y C A L E N D A R	?	0	3	0	J A N	0	3	0	J U L	
	N	84	779	2		65	798	1		
	Y	1	0	1		0	1	2		
	?	0	3	0	F E B	0	3	0	A U G	
	N	73	789	3		89	774	1		
	Y	1	0	1		0	1	2		
	?	0	3	0	M A R	0	3	0	S E P	
	N	59	803	3		98	766	1		
	Y	1	0	1		0	0	2		
	?	0	3	0	A P R	1	2	0	O C T	
	N	46	816	3		264	601	0		
	Y	1	0	1		0	0	2		
?	0	3	0	M A Y	2	1	0	N O V		
N	26	837	2		461	404	0			
Y	0	0	2		0	0	2			
?	0	3	0	J U N	3	0	0	D E C		
N	44	818	2		650	215	0			
Y	0	1	2		0	0	2			

Data table for Figure 11:

**Employment at a Job/Business in Calendar Year 2007 as Reported in SIPP and in the EHC
(Y=yes; N=no; ?=missing), by Month, Collapsed Across Site**

Note: N=1,620. Missing data (shaded cells) are ignored in the statistical analysis.

		SIPP				SIPP			
		?	N	Y		?	N	Y	
E V E N T H I S T O R Y C A L E N D A R	?	1	9	7	J	2	10	5	J
	N	60	463	70	A	29	488	82	U
	Y	104	37	869	N	88	33	883	L
	?	1	8	8	F	4	8	5	A
	N	57	470	61	E	41	480	62	U
	Y	92	36	887	B	105	35	880	G
	?	1	7	9	M	4	8	5	S
	N	50	472	67	A	41	483	60	E
	Y	76	37	901	R	117	41	861	P
	?	0	8	9	A	7	7	3	O
	N	35	482	72	P	176	367	38	C
	Y	68	37	909	R	325	30	667	T
?	0	11	6	M	8	6	3	N	
N	28	487	74	A	307	253	20	O	
Y	51	34	929	Y	566	19	438	V	
?	1	11	5	J	15	2	0	D	
N	28	488	83	U	448	131	6	E	
Y	61	39	904	N	804	10	204	C	

Data table for Figure 12:

School Enrollment in Calendar Year 2007 as Reported in SIPP and in the EHC (Y=yes; N=no; ?=missing), by Month, Collapsed Across Site										
Note: N=1,620. Missing data (shaded cells) are ignored in the statistical analysis.										
		SIPP					SIPP			
		?	N	Y			?	N	Y	
E V E N T H I S T O R Y C A L E N D A R	?	2	9	2	J A N	?	1	12	0	J U L
	N	101	1255	31		87	1371	29		
	Y	40	13	167		11	59	50		
	?	3	9	1	F E B	?	1	12	0	A U G
	N	83	1271	31		106	1269	34		
	Y	40	17	165		26	39	133		
	?	3	9	1	M A R	?	0	13	0	S E P
	N	66	1294	27		123	1246	20		
	Y	36	16	168		19	25	174		
	?	1	11	1	A P R	?	4	9	0	O C T
	N	47	1311	30		435	938	15		
	Y	34	18	167		57	18	144		
?	2	10	1	M A Y	?	6	7	0	N O V	
N	29	1332	29		755	625	8			
Y	28	21	168		113	12	94			
?	1	12	0	J U N	?	9	4	0	D E C	
N	54	1381	37		1091	296	3			
Y	16	48	71		162	6	49			

Appendix 4: GENMOD MODELING DETAILS

The general approach to the modeling analyses summarized in this paper was very simple in that it did not use any individual weights, and thus ignored key departures of the SIPP design from a simple random sample. A very practical reason for this decision was the fact that the weights were not available until quite recently, much too late to meet our goal of assessing the results of the small scale pilot test quickly in order to make decisions about the future direction of the re-engineering effort. More fundamentally, however, we viewed the analysis as a basic test of the new methods among the set of people who responded to both surveys, rather than an exercise to produce estimates for some larger population.

Thus, the GENMOD modeling exercise in effect assumed a simple random sample of individuals. However, through the Generalized Estimating Equations (GEE) procedures, it did allow for the correlation of responses within individuals, and thus accounted for the fact that repeated measurements were derived from the same individual.

Our approach was exploratory. All models started with the set of SAS statements shown below:

```
proc genmod descending data=[filename];  
class personid month site method;  
repeated subject=personid / type=exch;
```

We followed those statements with various model specifications to assess a “main effect” for METHOD:

```
model response = month|site method / dist=bin noint;
```

METHOD X MONTH interactions:

```
model response=month method*month / dist=bin noint;
```

and interactions with the SITE factor:

```
model response=month month*method site month*site method*site / dist=bin noint;
```

This generic approach was the starting point for all characteristics examined in this report. In some cases, the initial results led us to develop and run additional models to further explicate the details of a particular finding or effect.