

Evolution and Change in the Consumer Expenditure Surveys: Adapting Methodologies to Meet Changing Needs

Karen Goldenberg
Jay Ryan

U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions
Division of Consumer Expenditure Surveys

National Bureau of Economic Research, Summer Institute 2009
Conference on Research on Income and Wealth
Cambridge, Massachusetts

July 13, 2009

This document reflects the opinions of the authors and does not constitute official policy for the Bureau of Labor Statistics. The authors gratefully acknowledge the contributions of Jennifer Edgar, Scott Fricker, Jeffrey Gonzalez, Steven Henderson, Geoffrey Paulin, Carolyn Pickering, John Rogers, Mark Vendemia, and Nick Zwelineff. The authors take full responsibility for any errors.

Table of Contents

1.	Introduction.....	1
1.1	BLS Mission	1
1.2	CE Background.....	2
1.3	Rationale for Survey Evolution and Change	3
2.	Evolution and Change in the CE Production Surveys	5
2.1	Interview Survey: Major innovations	5
2.2	Interview Survey: Systematic Revisions	9
2.3	Diary: Major innovations.....	12
2.4	Interview and Diary Surveys	13
2.5	CE Data Dissemination.....	16
3.	Research Completed, Not Implemented	19
3.1	Respondent Incentives	19
3.2	Individual diaries	20
4.	Current and ongoing studies	22
4.1	Alternative Approaches for the CE Interview Survey	22
4.2	Feasibility test: Information Book for Telephone Respondents	23
4.3	Bounding.....	24
4.4	Validation/Qualitative Exploratory Study	24
4.5	Order effects.....	25
4.6	CE Surveys Redesign: The Gemini Project.....	25
5.	Planned or proposed research	27
6.	References.....	28
	Appendix A: Major Revisions to the CE Interview Survey, 2004 through 2009 and Proposed for 2011.....	31
	Appendix B: CE Source Selection.....	34

Evolution and Change in the Consumer Expenditure Surveys: Adapting Methodologies to Meet Changing Needs

Karen Goldenberg and Jay Ryan

1. Introduction

On July 22-23, 2002, the National Bureau of Economic Research-Conference on Research on Income and Wealth (NBER-CRIW) sponsored a Consumer Expenditure Survey (CE) Research Conference (<http://www.nber.org/~confer/2002/si2002/cexprg.html>). This paper highlights some of the work accomplished by the CE program in the seven years since that conference, and discusses the CE program's plans for future survey redesign and improvement. Although the basic *design* of the survey has been in place for nearly 30 years, both the content and the methods governing survey operation—data collection, data editing and processing, estimation, and analysis—have changed considerably in the intervening years. The changes are not necessarily obvious, and in many cases have been constrained by the need to maintain continuity and consistency for the Consumer Price Index (CPI). The paper will address:

- Section 2. Questionnaire and methodology changes implemented into the ongoing production surveys that resulted from internal research. These changes include systematic updates to the surveys every two years, and have been motivated by issues of falling response rates, response burden, measurement error, and new data collection technologies.
- Section 3. The CE research that has not been implemented in production or tested in a field environment, either due to a lack of resources or because the results did not meet expectations
- Section 4. Current and on-going studies
- Section 5. Future research directions, planning for and leading to a redesign of the CE surveys

1.1 Bureau of Labor Statistics (BLS) Mission

The BLS Mission Statement states that, "BLS data must satisfy a number of criteria, including relevance to current social and economic issues, timeliness in reflecting today's rapidly changing economic conditions, accuracy and consistently high statistical quality, and impartiality in both subject matter and presentation." (<http://www.psb.bls.gov/whatisbls/strategic/mission.php>) The CE program supports this mission in its research and development projects. In particular the CE program maintains the relevance of the surveys through research projects and regular biennial updates to the Interview Survey questionnaire, and improves data quality through methodological and data collection revisions. A staff of economists, survey methodologists, and statisticians work together to identify and propose solutions for problems of measurement error, to develop new approaches to maintaining response rates and other data collection issues, to streamline data processing procedures, and to investigate and implement more efficient estimation methods.

The CE measures spending by consumers for the total United States noninstitutional population. The principal mission of the survey is the biennial update of the relative weights of goods and services for the CPI, the primary measure of inflation, and one of the most important economic measures of the U.S. economy. Because the expenditure estimates are a critical input into the CPI, it is vital to have accurate information on consumer spending habits. In addition to expenditure data, the CE collects information on the amount and sources of family income, housing characteristics, and detailed demographic information, making it a unique data source for policy analysis and research.

1.2 CE Background

The CE program consists of two independent surveys, the Diary and the Interview Survey. The samples for the surveys are updated every ten years based on the most recent Decennial Census, along with additions for new construction. The unit of measure is a consumer unit (CU), which consists of all members of a household who are related or who share expenditure decisions. The data are collected by the Census Bureau for BLS.

In the Diary Survey, respondents complete two one-week expenditure diaries that capture all expenditures for everyone in the CU. The diaries span two consecutive weeks. The Diary Survey is designed primarily to collect expenditure data on small or frequently purchased items such as food, meals away from home, apparel, and personal care items. Interviewers visit the selected sample units to collect demographic and income information, leave the diary, and ask the diary-keeper to record expenditures daily. The Diary Survey is based on a sample of about 7,000 CUs per year.

The goal of the Interview Survey is to capture expenditures for larger and less frequently purchased items, and regular payments, such as rent or utilities, that occur monthly. The Interview Survey is a quarterly survey with a rotating panel design. Each CU is interviewed five times over the course of 13 months. In the first interview, the interviewer collects demographic data for all CU members and establishes an inventory of items such as housing, vehicles, and insurance policies, which are updated in subsequent interviews. The expenditure data collected in the first interview are used for "bounding" purposes only, ensuring that reported expenditures are not duplicated in interview two. In interviews two through five the interviewer collects data on all expenditures for the previous three month period, and updates demographic and inventory data. Income and assets data are collected in interviews two and five only. The Census Bureau completes approximately 35,000 CE interviews per year.

While there is some overlap in spending categories covered in the two surveys, some items, such as personal care items, are not captured in the Interview Survey while others, such as food expenditures, are captured only at a global level in the Interview Survey.

The revision of the CPI remains the primary reason for undertaking the CE. However, CE data are used by a wide variety of researchers, policy analysts, and government agencies who are interested in studying the economic well-being of American families. One importance of the CE to external users is that it allows data users to relate the expenditures and income of consumers to the characteristics of those consumers. The CE data are used by government agencies,

researchers, and economic policymakers interested in the effects of tax or other policy changes on levels of well-being among diverse socioeconomic groups. Econometricians use the CE data to construct economic models, for example, to estimate structural models of optimal life-cycle consumption expenditures, to test whether people spend based on their current income or on what they expect to earn over a longer period of time, or to investigate the distributional effect of consumption taxes. Market researchers find CE data valuable in analyzing the demand for groups of goods and services. U.S. Government agencies that draw on CE data include:

- Bureau of Economic Analysis (BEA) - uses the survey data to compute implicit rents for homeowners in the Personal Consumption Expenditures; as a source of information for revising its benchmark estimates of nursery schools and day care; and in calculating the Group Health Insurance component of Employer Contributions for Employee Pension and Insurance Funds in the expenditure and income components of the National Accounts.
- U.S. Department of Defense - uses CE data to calculate the spendable income table and the continental United States Cost of Living Allowances (COLA). The COLA compensates for a portion of excess costs for non-housing expenses incurred in areas that exceed costs in an average U.S. military location by more than 8 percent.
- Internal Revenue Service (IRS) - uses special tabulations produced by the CE on the purchase of taxable items to calculate the standard state and local general sales tax deductions (published yearly in the IRS 1040 Instructions),
- U.S. Census Bureau - uses CE data to produce annual Alternative Poverty Estimates in the United States, as recommended by the National Academy of Sciences. The New York City government also produces alternate poverty estimate using CE data, and uses these measures in guiding a set of anti-poverty initiatives.
- Office of Personnel Management (OPM) - uses a special CE Income table prepared annually. OPM uses the data as weights in calculating the COLA to white-collar civilian Federal employees in Alaska, Hawaii, Guam and the Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands.

1.3 Rationale for Survey Evolution and Change

The CE Survey was designed in the late 1970s as a paper and pencil survey collected in person. The overall design and structure of the questionnaire have remained essentially the same for the last 30 years, although both the survey content and the methods of data collection have changed. Since the survey was first designed, the technology for conducting surveys, knowledge about survey methods, and the overall social environment, data collection environment, and consumer environment have changed significantly. New challenges in the social environment include an increase in the number of single person households and the number of single-parent families,¹ a general perception that people feel rushed and lack free time,² and an increasing likelihood that

¹ The number of single person households rose from 22.7 percent in 1980 to 27.5 percent in 2008, and the number of single-parent families increased from 21.5 percent to 32.6 percent of families in that same time period (U.S. Census Bureau, 1993; 2008).

² Respondents to a Pew Research Center survey report that having free time is a top priority in their lives, but those who value it the most are the demographic groups least likely to have it—the employed, the middle aged, and mothers of young children. Other surveys document a growing perception of feeling rushed, a perception that parallels the growth of two-earner couples and employed mothers (Taylor et al., 2008).

the occupants of a sample unit will speak a language other than English. New challenges in the data collection area include a dramatic increase in the number of competing surveys of all types; less willingness on the part of households to share information/more privacy concerns; more public awareness of threats to privacy and confidentiality (with regular leaks), leading to an increase in the need for securing data; safeguarding respondent confidentiality; and escalating fuel costs, leading to and the increasing cost of visiting households in person. New challenges in the consumer environment include the expansion of on-line shopping, automatic bill paying, and the growth of large general merchandise stores, all of which lead to difficulties in respondent recall and reporting. As a result of these developments, response rates have decreased; underreporting has increased; and survey collection costs have increased.

One thing that has not changed is that the Interview Survey and the Diary Survey present burdensome and difficult tasks, both for respondents and for interviewers. Pressure to reduce burden competes with regular demands for new data. The use of new technology such as the Internet, electronic scanners, electronic data sets, and improved survey methodology is essential for the CE to respond to these issues in the data collection and consumer environments. To the extent possible, the CE program has conducted and will continue to conduct research to ease these burdens, but has been hampered to some extent by limited resources. The CE is a small program. It represents approximately 14 percent of the staff of the Office of Prices and Living Conditions (OPLC), or 10 percent of program-supported Federal and contractor OPLC staff when survey processing personnel are included. (For comparison purposes, the CPI figures are 32 and 39 percent, respectively. All percentages exclude the Field Economists responsible for repricing items in the CPI market basket.)

A continuing challenge to CE, to BLS, and to all Federal programs is that the Congress frequently adjourns a session without passing a budget for the next fiscal year. When agencies operate under a Continuing Resolution (CR), spending under the CR is limited to the level from the prior year's budget. In fiscal year 2008 BLS was funded \$30 million below the request in the President's Budget, a level that reverberated into the CR for the first half of fiscal 2009. Within BLS, spending under the CR has been limited to activities directly associated with conducting production surveys. These restrictions have placed further challenges on planning for and conducting new initiatives to respond to the challenges noted above.

2. Evolution and Change in the CE Production Surveys

This section presents major changes that have been introduced into the production surveys in the years since the 2002 NBER Conference. Many of the changes are not visible to users of aggregate level variables, as they occur in detail level questions, sometimes with but often without creation of new expenditure category codes for analysis. The changes include efforts to reduce response burden by designing data collection instruments that are easier for respondents to complete, and by phrasing questions so they are easier to understand and more accurately convey the intent of the question. Changes also address survey response through new field procedures and training materials.

2.1 Interview Survey: Major innovations

2.1.1 Automating the Interview Survey Instrument: Conversion to CAPI

In April 2003 the Interview Survey moved from a paper and pencil data collection form to a computer-assisted personal interview (CAPI) data collection instrument.³ The CAPI development project was an interagency effort with the Census Bureau that started in the late 1990s. It was the largest survey improvement project ever undertaken by the CE program, and was one of the first major government surveys to use CAPI data collection with the Blaise software. (Blaise provides a graphical user interface, while other software applications were MS-DOS based.) Activities in the conversion process included developing and testing the survey instrument and case management system (all phases from design specifications through functionality and usability testing); developing new data transfer systems; conducting iterative testing of all subsystems at various stages; and training interviewers, many of whom lacked computer skills. The development process culminated with a large scale "dress rehearsal" field test in 2002. The dress rehearsal, which lasted 9 months and involved all 12 Census Regional Offices, was deemed successful and the decision was made to move the production survey to CAPI (Henderson and Swanson, 2003).

CAPI collection began in the production survey in April 2003. All data collected on paper questionnaires in the first quarter of 2003 were converted to electronic form and transitioned to the CAPI instrument at the start of production. That way, all CUs were interviewed by CAPI starting in April even if their earlier interviews had been on paper.

Automated data collection brings with it many advantages over paper and pencil. These include:

- Automating and enforcing complex skip patterns. The interview contains numerous screening questions with detailed follow-up questions or skips to different sections of the questionnaire. CAPI automatically sends the instrument to the correct place, and ensures that the interviewer does not miss any sections. Under paper collection, questions or even entire sections of the questionnaire were sometimes overlooked, requiring post-collection clerical review.
- Interviews are shorter. The instrument automatically computes elapsed time for each interview section, leading to an accurate assessment of the length of the interview. The

³ This discussion has been extracted from L. Groves (2003).

average interview in 2009 is 65 minutes. Estimates for the paper form were approximately 90 minutes, although exact data were not available.

- The instrument includes built-in range checks on reported dollar amounts. Interviewers must confirm or correct extremely low or extremely high values, and are encouraged to add a note to any field to explain any unusual expenses. This information helps CE staff during the review process.
- The instrument identifies and allows the user to delete duplicate expenditures that are reported in both the current and previous interviews in those instances where the recall periods overlap.
- The instrument collects "paradata," i.e., data about survey operations, such as the length of time each section takes or the number of contact attempts required before resolving a case (see section 2.4.2).
- CAPI eliminates the need for clerical transcription of data for use in the next interview, as well as all clerical keying and coding operations. As a result, data are cleaner and available to BLS earlier than had been true with paper forms.
- The CAPI instrument contains several interviewer assistance resources, including online help screens with examples of the type of expenditures for each item, Frequently Asked Question (FAQ) screens to help the interviewer address respondent questions and concerns, and an online searchable product database which allows the interviewer to enter an item description and determine where that item should be reported.
- CAPI provides survey managers more timely data on the status of cases. Field supervisors can obtain reports for monitoring interviewer performance based on paradata about how the survey was conducted.
- The Blaise software captures individual key strokes made during the interview, and CE economists can use this information during data review. This "audit trail" data can alert reviewers to problems with data collection in a specific case and may also provide them with more general insights into usability or problems navigating the CAPI instrument.

In principle, another benefit of CAPI is to permit faster changes to the collection instrument, making it easier for the survey to stay relevant to the changing economy. The CE was able to add questions to the Interview Survey about Medicare prescription drug expenses in time for the start of this program in 2006, and modified the instrument in 2008 to collect data on economic stimulus tax rebate payments. However, modifying the CAPI instrument is not as fast or as easy as developers had anticipated. All changes require extensive testing to ensure that the instrument captures the intended data, and this testing involves a significant amount of time and staff resources.

Overall, the conversion to CAPI has been very beneficial for the program—in terms of data quality, survey operations, timeliness, and the additional data now available about the interview process. If there is a disadvantage, it is that staff have become less knowledgeable about the content of the questionnaire. The lack of a physical paper form makes it more difficult for new employees to learn the many sections, questions, and skip patterns of the questionnaire. Another, less significant, drawback is that it is more difficult to identify certain types of calculation errors that were obvious on the paper forms.

2.1.2 Redesigned Information Book for Respondents

The CE Information Book is a spiral-bound book with lists of relevant items for most of the Interview Survey sections. Interviewers show it to respondents to provide "cues," as cues are more effective than "unaided recall" in retrieving information from memory (Bradburn and Sudman, 1991; Tourangeau, Rips, and Rasinski, 2000). The CAPI instrument and the Information Book have been designed to work together, with the relevant page numbers for a section appearing on each screen in the CAPI instrument. Interviewers are trained to preface their questions with "Looking at pages X to Y of the Information Book," and ask respondents to look over the examples for a section. Research by CE staff supports the theoretical premise; it demonstrates that when respondents actually look at the Information Book (Edgar, 2006), they are more likely to report items that they purchased, and to report them in the correct interview category (as defined by CPI).

While the CE program recognized the importance of the Information Book, interviewers reported that it was difficult to get respondents to pay attention to it. Generally, interviewers receive one copy, which they carry with them to each interview. Interviewers complained that the book wasn't attractive, and that the paper ripped and tore away from the spiral binding, making it even less interesting to respondents. In 2007, the CE program revised the book to make it more visually appealing and relevant. It is now printed in color, on laminated pages that are much less likely to wear out, with spiral bindings that lie flat and allow for easy page turning. The photographs on the front demonstrate a broad range of expenditures, and the book incorporates plenty of "white space" around the text to make it easier for respondents to scan and recognize purchases.

2.1.3 New Field Procedures for Telephone Interviewing

As noted above, the Interview Survey was designed for personal visit collection, with no provision for interviewing by telephone. However, CE management was aware that some telephone interviewing was taking place. The move to CAPI collection gave the program the ability to monitor collection mode for the first time. Mode of data collection is one of four main sources of measurement error—along with the data collection instrument, the respondent, and the interviewer (Groves, 1989). As defined by Groves, measurement error is the difference between the true value for a measure and the value obtained by the survey. Measurement error due to mode is the difference between the report that would be obtained in person and the one obtained over the telephone. In theory, since the survey was designed to be conducted in personal visit face-to-face interviews only, measurement error from mode (or "mode effects") should not be an issue.

Initial research on mode showed that a surprising 40 percent of interviews were conducted by telephone (McGrath, 2005), and that there were systematic differences by mode. Telephone use varied widely by Regional Office, and some interviewers completed very large proportions of their cases by phone. In addition, respondents differed by collection mode. Telephone respondents had higher incomes than face-to-face respondents, but they were less likely to provide complete income data. While telephone respondents reported almost the same level of expenditures as personal-visit respondents, they did so in interviews that were roughly 25 percent shorter. Since longer interviews are viewed as "better" for reporting expenditures, there was concern that phone respondents were underreporting their purchases. In addition, telephone

respondents' answers were much more likely to need data adjustment in the form of imputation (because no costs were reported) or allocation (because the respondent provided one cost that covered several purchased items).

Safir and Goldenberg (2008) subsequently explored the mode effect and circumstances surrounding use of the telephone for CE interviewing in considerable depth. Their data showed telephone interviewing at 36 percent, based on four years of CAPI data rather than McGrath's one year. In looking for explanations, they discovered that the Interviewer's Manual had relatively little to say about telephone interviewing. However, the percentage of telephone interviews suggests that interviewers were using the phone for more than a "last resort." On the other hand, the question measuring use of the telephone was somewhat ambiguous, asking, "How did you collect most of the data for this case?" It wasn't clear if "case" meant "this interview" or "all visits to this household," and "most of the data" is open to a variety of interpretations.⁴

In response to these findings, CE staff took two actions. First, they developed a set of procedures that specify *when* telephone interviewing is allowed, introducing a requirement for supervisory approval for any deviations from the guidelines. These procedures emphasize the fact that CE is supposed to be a personal visit survey, and require interviewers to make every effort to conduct at least the first interview in person. The guidelines also include instructions on *how* to conduct an interview by telephone, because the proliferation of cell phones means that interviewers must be sensitive to the fact that a respondent may not be at a convenient or appropriate place at the time an interviewer calls. The CE staff anticipate that the new guidelines—to go into effect in August, 2009—will standardize the use of telephone interviewing across interviewers, supervisors, and Regional Offices.

Second, CE staff developed a new series of paradata items in the CAPI instrument to better measure the use of telephone interviewing. These questions were introduced with the April 2009 interview. They differentiate between complete and partial use of the telephone and ask interviewers to identify the sections that were collected entirely by phone.

2.1.4 Nonresponse Bias Study

In 2006, the Office of Management and Budget (OMB) issued a directive for any household survey with a response rate below 80 percent, which includes the Interview Survey. The directive requires an analysis of nonresponse to determine if the data are missing completely at random (MCAR), and another analysis to estimate the amount of nonresponse bias in the survey's estimates. The CE staff responded to the OMB directive by conducting four different studies, each of which took a different approach to looking at nonresponse bias. The four studies are:

- A comparison of response rates between subgroups of the survey's sample;

⁴ Safir and Goldenberg (2008) looked at detailed expenditure data, while McGrath (2005) focused on summary-level variables. While the McGrath research suggested that the telephone was the main reason for differences in expenditures, Safir and Goldenberg (2008) demonstrated that most important difference was due to behaviors that are facilitated by an in-person interview, especially the use of records and the Information Book. When these were controlled, the differences between telephone and face-to-face interviews largely disappeared.

- A comparison of respondent demographic characteristics to those of the American Community Survey (ACS);
- An analysis of nonresponse bias using ‘harder-to-contact’ respondents as proxies for nonrespondents; and
- An analysis of nonresponse bias using intermittent respondents and attriters (survey dropouts) as proxies for nonrespondents.

The four studies were designed to answer three questions: (1) Are the data in the Interview Survey MCAR? (2) What are the demographic characteristics of the nonrespondents? and (3) What is the level of nonresponse bias in the Interview Survey?

All four studies addressed the MCAR question, and concluded that the data in the Interview Survey are not MCAR. Characteristics of nonrespondents were examined using frame variables, ACS data, and proxy nonrespondents. Collectively, these studies showed that blacks are underrepresented among the respondents, and those over 55 years old are overrepresented.

Estimates of nonresponse bias could only be calculated in two of the four studies. These studies identified subgroups of the survey’s sample that were thought to exhibit the same behavior and characteristics as nonrespondents, and then treated them as ‘proxies’ for nonrespondents. Using the mean expenditure estimates from the survey’s respondents and these proxy nonrespondents, nonresponse bias was computed for total expenditures as well as thirteen expenditure subcategories, ranging from regular monthly expenses such as housing payments to infrequent and highly variable expenses such as those for education. Neither study showed evidence of nonresponse bias for total expenditures. For the thirteen subcategories, there was little meaningful or consistent bias found across the studies. As a result, the Interview Survey seems to provide a counterexample to the commonly held belief that if a survey’s missing data are not MCAR then its estimates are subject to nonresponse bias. The study concluded that nonresponse bias does not appear to be a significant issue for the Interview Survey.

2.2 Interview Survey: Systematic Revisions

The CE program updates and revises the Interview Survey every two years, in conjunction with the CPI market basket revision. The revision process begins shortly after the last set of biennial revisions goes into production, and is the mechanism by which the survey stays abreast of changes in the consumer environment. Research into potential changes includes:

- A review of new or emerging products and services that are not currently collected
- Identification of changes in the ways existing products or services are marketed and billed (e.g., bundled utilities)
- A review of selected expenditure categories to determine whether products need their own category codes
- A review of questionnaire content to identify and correct any items that have been identified as problematic or confusing for respondents, or difficult for interviewers to handle within the instrument
- Soliciting input from CPI and other BLS users, Field Staff, and others about needed or desired changes.

CE staff review all proposed changes and determine which of the proposals to implement. Items that require wording changes are generally easier to introduce than proposals that require significant programming and processing modifications, both at Census and at BLS. Once new questions have been proposed, they are tested in the BLS Cognitive Laboratory (Section 2.4.1) to ensure that the revised wording is understood by respondents, and that it captures the specific item(s) intended but not others that are outside the scope of that topic.

The revision process also involves preparation of software specifications, programming, and testing the new questions and the revised instrument. In addition, the field staff develops training materials to introduce the new or changed questions to interviewers, and conducts interviewer training prior to implementation.

CE implemented major revisions in 2005, 2007, and 2009, and more are planned for 2011. In addition to the biennial revisions, CE sometimes introduces special-purpose changes on years between the scheduled revisions. Minor off-year changes such as questions to capture the 2008 economic stimulus payments occur much more quickly than the regular two-year cycle changes. Because of the survey's three-month reference period, all biennial changes are introduced in April of the relevant year, allowing collection of expenditures for a calendar year.

Table 1 contains selected examples of changes implemented since the conversion to CAPI. Appendix A contains the full list of updates.

Table 1. Selected Examples of Revisions to CE Questionnaire Items

Year Change Implemented	Description of Changes
2004	<ul style="list-style-type: none"> • Deleted the quantity consumed and the unit of measure questions for utilities • Added a separate question for money put into educational savings accounts (e.g., 529 plans)
2005	<ul style="list-style-type: none"> • Updated the types of “technology” items and services collected such as television sets, DVD players, personal music player (e.g., Ipods, PDAs), and fees for services that use Global Positioning Service (GPS) (e.g., OnStar) • Streamlined questions for collecting home equity loan payments • Added references to online and automated payments as cues to respondents to help them remember purchases made this way, particularly for services • Created a new section to collect internet services such as downloading audio or video files and online gaming
2006	<ul style="list-style-type: none"> • Created a new section to collect data on CU member enrollment in the Medicare Prescription Drug plan, Medicare Part D
2007	<ul style="list-style-type: none"> • Improved the collection of data on mortgages, including interest only and variable rate mortgages • Started collection of the property value of owned properties every quarter instead of only once • Added a specific code for the collection of Voice Over Internet Protocol (VOIP) services in the telephone section • Redesigned the structures of the telephone and utilities sections to reduce the amount of time to collect the data in these sections and “double reporting” of data from quarter to quarter; also improved the collection of combined services items • Updated collection of video and computer game hardware and software • Added questions about reverse mortgages • Improved question wording throughout the survey
2008	<ul style="list-style-type: none"> • Added questions to capture the receipt, amount, and primary use of the 2008 economic stimulus payments
2009	<ul style="list-style-type: none"> • Created a separate item code for "test preparation or tutoring services." • Added questions to capture the receipt and use of the one-time \$250 stimulus payment for Social Security recipients and other retirees • Created a separate item code for expenditures on "portable media." • Moved the item codes for "Care for invalids, convalescents, handicapped, or elderly persons in the home" and for "Adult day care centers" from Section 19 to Section 15 so as to capture reimbursements for these items.
Proposed 2011 (Subject to change)	<ul style="list-style-type: none"> • Further improvements to collection of rental equivalence by rewording the questions related to timeshares and vacation homes available for rent; also by better identifying rental units contained within a sample unit such as a basement apartment • Add new questions to support alternative poverty thresholds, such as more detail on school meals, subsidized utilities, and the Earned Income Tax Credit • Improve collection of bundled technology services (e.g., telephone, internet, and television) and equipment such as phones and satellite receivers and modems • Reduce respondent burden when reporting health insurance premiums by calculating, rather than re-asking, premium amounts • Streamline the household repairs and trip sections • Introduce new products and services

2.3 Diary: Major innovations

2.3.1 The Respondent-Friendly Diary

The design of the CE diary form has been reviewed and revised over the years, but continued to be a challenging document for respondents. The CE program began research in 2000 on improving the diary document to make the respondent's data entry task easier (Davis, Stinson, and To, 2002). The goal of the diary redesign project was to improve response rates while maintaining the same level of data quality. Steps in this process included simplifying the form layout, improving navigation throughout the form, and reducing the number of major categories. Various prototype forms were tested and refined. One redesign version was selected for a major field test, which was conducted in the fall of 2002.

The successful field test of the new form led to some additional modifications in that form, which was implemented in production in January 2005. The redesigned form is based on a standard 8-1/2 by 11 inch page, portrait rather than landscape orientation, and is more visually appealing. It incorporates color and photographs, clearer instructions, and more effective "sample" pages than the predecessor version. Checkboxes improve item classification and make it easier for respondents to record their expenses. The diary contains FAQs for respondents, a pocket for receipts, and tabs that fold over the page so that instructions and categories are visible while a respondent is recording data.

To support the redesign project, the Census Bureau's National Processing Center introduced a new keying and item coding system. Where possible, the system assigns the item code based on the expenditure description recorded by the respondent. If a code cannot be assigned automatically, the system displays a list of likely expenditure categories and the keyer makes a selection. Data are keyed twice. When the item code values assigned by the two keyers differs, a keying adjudicator determines the item code.

With the new diary in place, the response rate improved, and data quality measures were generally positive, although allocation rates for some items were higher as a result of many broad categories being removed. On the one hand, because of the new checkboxes, allocation rates in the "food away from home" section fell from 34.8 percent to 5.9 percent. On the other hand, the allocation of "food at home" items increased from 7.8 percent to 9.9 percent due to the open-ended recording of expenses in that section. CE staff expected the open-ended format to result in less detail, but accepted that drawback as a trade-off for simplifying the recording of expenditures (To and Tan, 2008).

2.3.2 CAPI Collection of Demographic and Household Data

During diary placement, the interviewer collects demographic data for the household. The interviewer also obtains CU income information at the time the second week diary is collected. Following the conversion of the Interview Survey to CAPI, the interview portion of the Diary Survey was also converted to CAPI. It went into use in March 2004.

2.4 Interview and Diary Surveys

2.4.1 Asking Better Questions

The Cognitive Aspects of Survey Methodology movement, beginning in the 1980s, was one of the most significant innovations in the emerging field of survey methodology. An underlying premise of the movement is that the principles of cognitive psychology can be applied to the basic building-block of a survey, the question and the response. By understanding the mental processes involved when people store information in their memories, researchers can write questions to take advantage of those processes. This effort results in questions that respondents are more likely to interpret as the survey designer intended them, and that can be answered more readily and more accurately. When respondents interpret a question differently from the researcher's intent, and provide an answer based on that interpretation, the result is measurement error—in this case, a difference between the true answer for the intended survey question, and the answer provided by the respondent.

Survey designers, including CE staff, may write questions that are complex and hard to understand, or that present respondents with a lot of information to keep in mind while they formulate an answer. Such questions may incorporate a long list of items for the respondent to include (or not include) in the answer. However, research into the structure of questions and question wording suggests that sometimes the complexity added by the additional precision comes at a cost in respondent confusion or misunderstanding. Questionnaire design experts review questions and attempt to simplify and ensure that they are presented in the most effective way to the largest number of potential respondents to those questions.

The OMB requires that Federal agencies "ensure that all components of a survey function as intended when implemented in the full-scale survey and that measurement error is controlled by conducting a pretest of the survey components or by having successfully fielded the survey components on a previous occasion" (OMB, 2006:9). While most of the Interview Survey satisfies the requirement by virtue of being an ongoing survey, CE tests all new and revised questions in the Cognitive Laboratory operated by the BLS Office of Survey Methods Research. Laboratory staff maintain a pool of individuals from the local community who serve as respondents. Research staff develop testing protocols in which they ask respondents the survey questions, and then seek additional information to get a better understanding of what question the respondents answered (the intended one?) and how the respondents went about formulating their answers. Using a set of standard cognitive testing techniques, researchers identify any issues they discover about each question or question set, and propose solutions to address those issues. Revised questions are re-tested until it has been determined that the questions can be understood properly, answered accurately, and function as intended.

2.4.2 Contact History Instrument

The Contact History Instrument (CHI) is one tool in the collection of approaches being developed to understand and combat survey nonresponse. It is an automated standalone instrument that works in tandem with a CAPI survey instrument, and was designed to record information about every attempted and actual interaction between an interviewer and a CU. For each in-person or telephone contact, interviewers record (Piani, 2004):

- The nature of the contact, e.g., completed interview, no adult home, respondent too busy, language problem and the language used by the CU
- The nature of the noncontact, e.g., no one home, unable to enter gated community, telephoned/no answer
- Any issues or concerns perceived by the interviewer as being problematic for the respondent and that will affect response, e.g., too busy, not interested, doesn't trust government, confidentiality concerns
- The strategies the interviewer used to reach the CU, such as providing an advance letter, leaving a message on answering machine, or requesting a specific type of letter from the Regional Office.

Because it is an automated application, if the interviewer records data at the time of the contact, CHI automatically records the date and time of the entry. Otherwise, the interviewer enters the date and time.

CHI arose from a need identified by researchers seeking to understand and to reverse the decline in survey response rates. Groves and Couper (1998) developed a conceptual framework to explain the processes involved in contacting sample units and successfully persuading the residents of those sample units to participate in an interview. Their framework presents variables that are under the control of the survey designer, such as the survey topic and mode of administration, and those that are not, such as current economic conditions and the structure of the household. The framework also shows how these sets of variables operate, both directly and indirectly, on a respondent's decision to cooperate or refuse. Part of the Groves and Couper approach is based on training interviewers to respond appropriately to various types of householder situations and comments, and to find ways to draw on those comments so as to increase the likelihood of obtaining an interview. Data from CHI are critical for developing appropriate responses and strategies for addressing different types of situations.

Contact data offer a certain amount of predictive power as to which CUs will eventually respond. Data from the initial implementation of CHI in 2005 (Bates, 2006) show that CUs in the Interview Survey expressing no concerns about the survey were much less likely to be refusals (1 percent) than CUs that expressed concerns at every contact (36 percent). CHI data from earlier interviews help interviewers to prepare for contacts with CUs at the next interviewing wave, especially if there are issues that the interviewer can be ready to address (Henly and Bates, 2006). The CHI data may also help interviewers to plan their contact attempts by day and time of day for CUs with various characteristics. McGrath and Tan (2007) used CHI data to look at the likelihood of the first contact yielding a successful interview. They found that the best times for contacting respondents are not necessarily the best times for gaining cooperation. They also found that weekday daytimes work well for obtaining an interview, especially early weekday mornings (8:00 to 10:00 a.m.).

The CHI was implemented with the production Interview Survey instrument in April 2005, and in the Diary Survey in 2006.

2.4.3 Field Training Improvements

Interviewer training is a critical component of any survey that relies on interviewers to collect data. Well-trained interviewers are at the forefront of CE, responsible for maintaining response rates and contributing to high-quality data by ensuring that respondents participate actively and carefully in the data collection process. In the past few years, a Census-BLS Interagency Training Team has engaged in several efforts to evaluate and improve training for CE interviewers.

Several of these activities involved gathering information through focus groups with interviewers and field supervisors. These sessions were aimed at collecting specific information about the effectiveness of training methods and materials, as well as at providing members of the field staff an opportunity to offer their ideas and suggestions for improvements. The focus groups yielded a series of recommendations that were shared with both Census and CE management. After review, many of the recommendations were introduced into standard training practice.

Another major investment into the CE interviewer training program involved contacting over 50 government, private, and academic survey organizations to discuss their field training methods and evaluation procedures. This was a massive undertaking, but it allowed the Interagency Training Team to identify some ‘best practices’ used in other agencies, and to compare them with the current CE practices. Simultaneously with that effort, a contractor conducted a review of all CE training materials and spoke with several field staff. These two efforts resulted in an impressive list of recommendations. The recommendations were prioritized in terms of impact and feasibility, and items from the list have been and continue to be implemented. Several new training items have been created and distributed, including a Refusal Avoidance Job Aid and corresponding audio CD, and a “Selling the CE Survey” DVD. The team introduced new features to the training packages, such as adding learning objectives and learning checks for each chapter, adding more PowerPoint presentations, and incorporating more interactive training techniques. Finally, the team developed an interviewer evaluation form to be used for all training sessions, allowing both Census and CE to get continuous feedback on the training program as additional improvements are made.

2.4.4 Income imputation

Income is a key variable for the CE surveys, but it is also a sensitive item and one that is prone to item nonresponse. Prior to 2004, CE published income means for "complete income reporters." Respondents qualified as "complete income reporters" if they provided at least one value for a major source of income (wages and salaries, self-employment income, Social Security, or supplemental security income). Using this definition, even complete income reporters did not necessarily provide a full accounting of income.

Beginning with publication of the 2004 data, CE instituted a multiple imputation procedure to estimate values for income when respondents did not report precise values. Imputation allows the CE program to publish mean income that reflects all consumer units. For those who provide a bracket value for income, the imputed values lie within the range described by the bracket. When respondents provide no information about the value of income received, a regression-based method is used to derive the multiple imputations. Essentially, a regression is run to provide coefficients for use in estimating values for missing data points. The coefficients are

then “shocked” by adding random noise to each, and missing values are estimated using the shocked coefficients. To each of these estimated values, additional random noise is added, to ensure that consumer units (or members) with identical characteristics (e.g., urban service worker aged 25 to 34) will not receive identical estimates for their income. The resulting values are used to fill in invalid blanks where they occur, while reported values are retained. This process is then repeated four times, so that a total of five imputed values are computed for each missing value. In addition, for the small number of cases in which the respondent does not report receipt of any source of income collected at either the member or consumer unit level, receipt of each source is imputed using logistic regression. In each case where receipt is imputed, the income value is treated as a missing data point, and is imputed using the regression-based method described above. The method is based on Rubin (1987). See Paulin et al. (2006) for a full description, especially on using the imputed data in analysis.

2.4.5 Source Selection for Best Estimates

As noted above, the Interview Survey is designed to capture major expenditure items, household characteristics, and income data, and the Diary survey is designed to capture purchases that respondents are not likely to recall accurately over a long period of time. However, Diary respondents are asked to report *all* expenses (except overnight travel) that their CU incurs during the survey week. As a result, the Diary and Interview Surveys overlap considerably in their coverage of household expenditures. In a procedure known as integration, CE chooses the Diary or Interview component of the CE as the most statistically reliable source for each expenditure item, and uses data from that source for both CE publications and data comparisons. Appendix B provides a detailed explanation of the source selection process for the 2007 data.

2.4.6 Standard error tables

Sampling error is the difference between the survey estimate and the true population value. The most common measure of the magnitude of sampling error is the standard error. Beginning with year 2000 data, the CE program has made available standard error tables based on integrated data from both surveys. (<http://www.bls.gov/cex/csxstnderror.htm>) These standard error tables correspond to the program’s standard tables, except for the classification by region, population size of area of residence, and selected age, and are available on the CE Website. The primary purpose of standard errors is to provide users with a measure of the variability associated with the mean estimates. An article in the CE Anthology (<http://www.bls.gov/cex/anthology/csxanth5.pdf>, see Section 2.5.1 below) describes the use of standard errors in the CE. Also, BLS Regional Offices now use tests for significance when preparing their press releases on regional and metropolitan area spending patterns, and use the standard errors for this purpose when comparing expenditures across areas.

2.5 CE Data Dissemination

2.5.1 CE Anthology

The CE program has published three reports (2003, 2005, 2008) presenting articles discussing ongoing research and methodological issues pertaining to the CE and analytical articles using CE data. The reports include sections on survey research and methodology, on processing issues and improvements, and on analysis of topics of interest based on CE data.

(<http://www.bls.gov/cex/home.htm#publications>)

2.5.2 Microdata CDs

The BLS sells CD-ROMs containing annual CE public use microdata. The CD-ROMs contain either separate or various combinations of Interview Survey data, Diary Survey data, and tabulated data. A number of improvements have been made the last few years in the accompanying documentation, including the addition of a searchable ACCESS database of the data dictionary/codebook. Beginning in 2007, data formats were expanded to include an ASCII comma-delimited files, SPSS files, and STATA files. With the addition of the new formats, a STATA program to produce Diary Survey calendar year-adjusted averages and standard errors was added to the sample programs.

2.5.3 CE Microdata Users' Workshop

The first CE Microdata Users' Workshop was presented on July 17-18, 2006 in the BLS Conference Center. The agenda included two hands-on sessions using the microdata files in the computer training rooms in the conference center, with one session geared to inexperienced users and the other to more advanced users. General sessions covered topics related to the CE such as data processing and sample design, and specific topics related to the use of the microdata files, such as weighting and topcoding. Other sessions presented examples of how the CE microdata are being used in economic research. The fourth annual workshop is planned for July 29-31, 2009, and has been expanded to three days, including a full day of presentations by users on CE research topics. (<http://www.bls.gov/cex/csxworkshop.htm>)

2.5.4 New data classifications

In 2003, the CE program modified the questions on race and Hispanic origin to comply with new standards for maintaining, collecting, and presenting Federal data on race and ethnicity for Federal statistical agencies. In addition to these changes, a more comprehensive review was undertaken to evaluate the classifications of published CE data, with the goal of providing data users with additional information while maintaining reliability and continuity with previously published data. As a result of this review and the new race and ethnicity changes, a number of new classifications of data were made available, beginning with the publication of the 2003 data. Changes included:

- Publishing a separate table by Hispanic origin
- Adding a class for Asians in the race table
- Including additional higher income classes in the income table
- Including subclasses of homeowners for those with and without mortgages in the housing tenure table
- Including subclasses of urban consumer units for central city and for other urban areas in the type of area table

2.5.5 BLS and CE Website improvements

The BLS launched a completely redesigned website on July 21, 2008, making the website more accessible to data users. It is the world's highest traffic statistical website. As part of that redesign, the CE website added features such as navigation buttons which allow quick and easy access to CE reports, data tables, public-use microdata information, and a glossary of terms. In

the past few years, the CE has added new data and products to its website including, but not limited to, the following:

- Customer-friendly replication of the CE Computer Assisted Personal Interview (CAPI) instrument
- The CAPI Diary Household Characteristics questionnaire
- CE Anthologies
- The most recent biennial report
- Public-use microdata documentation
- Historical data tables back to 1984 in PDF and TXT format

Also available on the CE website is a one screen and multi-screen database (BLS LABSTAT) that allows users to obtain CE time series data using a form-based query application. A new feature allows for the creation of graphs from the selected data.

2.5.6 Visiting researcher program

The CE program participates in the BLS on-site visiting researcher program. More information on the program can be found on the BLS web site. (<http://www.bls.gov/bls/blsresda.htm>) Eligible researchers are given on-site access at the BLS national office to confidential data for approved statistical analysis. This is done within legal constraints to survey respondents to protect from disclosure individually identifiable data collected for exclusively statistical purposes under a pledge of confidentiality.

3. Research Completed, Not Implemented

3.1 Respondent Incentives

Falling survey response rates are a fact of life in the early 21st century, and CE's rates have fallen along with most other surveys. The survey literature consistently indicates that incentives have a positive effect on response rates, so CE attempted to combat falling response rates by conducting experiments with incentives. The experiment was successful in the Interview Survey but produced ambiguous results in the Diary Survey.

The experiments in both surveys provided prepaid unconditional incentives to half the sample in the production survey; the other half of the sample did not receive an incentive. One fourth of each survey sample received a \$20 debit card and one fourth received a \$40 debit card. The Census Bureau does not allow its interviewers to carry cash, but has successfully used debit cards as incentives in other surveys (Goldenberg, McGrath, and Tan, 2009).

3.1.1 Interview Survey

The Interview Survey experiment took place between November, 2005 and July, 2006. The incentive was distributed prior to the first interview; no incentives were paid after the first interview. Because of the Interview Survey's panel design, the last data were collected from incentive cases in July, 2007. In the analysis, results were reported separately for the first interview and aggregated across the second through fifth interviews, both because the incentive was administered at the first interview and because data from the first interview are not used in CE estimation.

At the first interview, the response rate for the \$40 incentive was almost five percentage points higher than that for the control group, which did not receive an incentive ($p < .05$). The effect of the \$40 incentive on response rates and refusals held through all five interviews, although the magnitude of the differences fell over time. The \$20 incentive was not statistically different from the no-incentive control group for most of the test. The incentive had little or no effect on the composition of the sample, but the \$40 incentive did show a trend toward more one-person households and urban respondents than the other groups.

Looking at data quality, there was little difference in reported expenditures with or without the incentive. However, respondents who received the \$40 debit cards in the first wave performed significantly better on the number of expenditure questions answered (more), the proportion of questions answered "don't know" or refused (fewer), the proportion who used records in reporting expenditures (more), and the percentage of reports that did *not* require allocation or imputation (fewer) (all $p < .05$). Recipients of the \$40 incentive were also generally more cooperative as reported in the CHI, in that they had fewer concerns about the survey, more respondents with "no concerns," and fewer "soft" refusals⁵ (all $p < .05$). Incentive households also showed a trend toward better income reporting (not significant).

⁵ A "soft refusal" is a situation where the interviewer contacts the respondent, the respondent does not complete the interview, and the interviewer codes the case as "reluctant respondent" in the CHI.

The \$40 incentive recipients were easier to contact; they required significantly fewer contact attempts than the control group, and resulted in a shorter field period between initial contact and case resolution (both $p < .05$) in the second through fifth waves. The \$40 debit card recipients were also more likely to cash the cards than recipients of the \$20 cards. While close to 80 percent of the recipients with completed interviews actually cashed their cards, only 17 percent of nonrespondents did so ($p < .05$). This is a positive outcome, given that Census recoups the cost of uncashed debit cards.

The CE program would like to introduce unconditional incentives as part of its regular data collection program in the Interview Survey. Recently obtained field cost data will help to determine if the savings in field operations could offset the cost of the debit cards. Without such a determination, there are no funds to support an incentives program (Goldenberg, McGrath, and Tan, 2009).

3.1.2 Diary Survey

The incentives experiment for the Diary Survey had one important difference from the test in the interview survey. In the Diary Survey, the incentive was directly related to consumer spending, but recording of that spending would begin shortly after the incentives were distributed, not three months later. The \$20 and \$40 debit cards were mailed to the incentive groups with the Diary Survey advance letter. There was some concern that the incentive might bias reported expenditures, but analysis indicated otherwise.

The Diary Survey experiment ran from March through November, 2006. The effect on response rates was minimal; both the \$20 and \$40 debit card rates were slightly more than one percentage point above the control group. Curiously, response rates for the Diary Survey *overall* were 3 to 4 percentage points higher during the incentive test period than they were before or after the test, an ambiguous result.

Incentive respondents did report more spending than control group respondents. In addition, CUs that cashed their debit cards reported more spending than either the control group or incentive recipients who did not cash their debit cards. Incentive respondents also performed better on most of the indirect data quality measures, such as having purchases recorded in the diary when the interviewer came to pick it up.

As in the Interview Survey experiment, effects on sample composition were small and not statistically significant (McGrath et al., 2007).

Because of the ambiguity of the test results, CE managers would like to repeat the experiment. Again, however, no funds are available to do so.

3.2 Individual diaries

The current Diary Survey design, in which one member of the CU records expenses in the Diary for all CU members, works to the extent that the main diary keeper is aware of other members' expenditures. However, this awareness assumes a high level of within-CU communication. At least one recent survey suggests that family interaction time has decreased, in parallel with an increase in internet use (Ortutay, 2009). In addition, in contemporary life the number of

"invisible" purchase options (e.g. music downloads, online gaming) has increased substantially. Individual diaries are one approach to remedying that situation, by having each member of the CU maintain a separate diary.

The CE contracted with the Census Bureau to develop and test individual diaries as a supplement to the main CU diary, conducted a small national feasibility study in 2006. The test raised both concerns and possibilities. The test had a low response rate (61 percent) when compared to the production sample during the same period (77 percent). On the other hand, for the sample units that did respond, individual diaries collected significantly more expenditures that might otherwise have gone unreported. Also, there was no evidence of duplicate reporting or of confusion as to which household member should be reporting specific purchases. Interviewers felt that the individual diaries emphasized the importance of within-CU communication and resulted in better reporting overall. The feasibility test was too small to produce national estimates, and was not designed for the purpose of analyzing expenditures. Nevertheless, compared to the production data collected at the same time, reported household spending was higher, not only because of the individual diaries, but in the main diary as well (Edgar et al., 2006).

CE management felt that the increase in reporting warranted further testing of individual diaries, and initiated activities to conduct a pilot test within the production survey. A joint BLS-Census Interagency Team began to plan the pilot test. A Steering Group developed charters for several teams, the two agencies were negotiating budget and cost agreements, and a general meeting took place in October, 2007 to formally start subteam activities. Unfortunately, when Congress ended the FY 2008 Continuing Resolution in early January, 2008, BLS received less funding than had been requested in the President's Budget, and CE was one of many programs that had to cut back on its activities. The pilot test was put on hold and has not been restarted.

4. Current and ongoing studies

4.1 Alternative Approaches for the Interview Survey

There is a widespread belief that Interview Survey data are underreported or exhibit other forms of measurement error (e.g., Tucker, 1985; Tucker, Meekins & Beimer, 2005). Underreporting has been variously attributed to recall error, panel conditioning, respondent fatigue, "satisficing behaviors" (i.e., making a minimal effort to answer the questions), and other causes. The length and perceived burden of the Interview Survey may also have deleterious effects on response rates.

The CE staff have been exploring alternative data collection strategies for the Interview Survey that could involve possible modifications to the survey instrument, the sampling techniques, and the estimation methods. The goal of investigating these design alternatives is to gain a better understanding of the feasibility and statistical consequences of implementing such changes in the current survey. Design alternatives under consideration have focused primarily on three methods: matrix sampling methods, adaptive (or responsive) designs, and imputation.

Matrix sampling methods, also called the split questionnaire design, generally involve dividing a lengthy questionnaire into sections of questions and administering each of these sections to different subsamples of the main sample. The goal of these methods, as suggested by Raghunathan and Grizzle (1995), is to reduce respondent burden (by offering a shorter questionnaire) and, in turn, possibly increasing the response rate. These methods have also been explored extensively in educational assessment and applied to well-known public health surveys such as the National Health and Nutrition Examination Survey (NHANES) (Gonzalez and Eltinge, 2007a; 2007b).

Adaptive designs, introduced by Groves and Heeringa (2004), suggest that survey design features can be adapted or tailored based on the information about sample members that is available prior to data collection and/or information that is collected as part of the data collection process. The goal of the responsive design is to use this accumulating data to improve the cost and error properties of the survey estimates.

Finally, the third area being investigated is the viability of imputation procedures to "fill in" gaps in the data that result because not all information is collected from every sample member under a matrix sampling design.

4.1.1 Split Questionnaire Test

The Split Questionnaire is the first such alternative that will be tested. The test will assess the effects of administering a shorter Interview Survey questionnaire on respondent burden, data quality, and nonresponse error. The study will consist of a three-wave panel design based on administration of a modified Interview Survey questionnaire to three groups in the first interview. The "full" questionnaire will actually be a partial Interview Survey instrument and will consist of 12 of the current 22 Interview Survey sections. For the second and third interviews, a control group will receive the full questionnaire again, but two treatment groups will each be administered a truncated or "split" questionnaire (in which the treatment groups will

each receive half of the full survey). Data from the split-questionnaire groups will be imputed in order to make response distribution comparisons to the control group.

This design will enable CE to perform data quality analyses using both direct measures (e.g., number of expenditure reports, expenditure amounts) and indirect measures (e.g., response rates, measures of perceived burden, item nonresponse, etc.), and to estimate nonresponse bias by comparing response rates, sample composition, and expenditure estimates across treatment conditions. The results from this study will be used to inform future research activities and decisions about how to redesign the production survey.

4.1.2 Statistical Support for Alternative Designs

Analyses to date have been simulation-based studies using historical CE data. These include:

- A modification to a simple matrix sampling design. Under this type of design, a sample unit would be naively assigned a subset of questions in an interview. In order to make a more informed assignment of questions and possibly improve the sampling error properties of the survey estimates, the approach borrowed ideas from the adaptive survey literature, and investigated the use of wave 1 expenditure information to inform the assignment of questions at wave 2. The process involved making simple adjustments to the base weights, which essentially amounted to dividing the base weight by the probability that a sample unit would receive a particular subset of questions, and comparing the variance due to the subsampling at the second wave. Estimates were made only using available cases, in other words, sample members for which data was directly collected (Gonzalez and Eltinge, 2008).
- Development of an imputation method for use in the estimation procedures to make estimates from all sample members and not just those directly observed. The research explores potential efficiency gains incurred from incorporating these methods into current weighting and estimation procedures (Gonzalez and Eltinge, 2009).

These simulations involve certain assumptions about the error structures of the current CE data. Future investigations will explore how these methods perform in the presence of nonsampling errors such as nonresponse and measurement error.

4.2 Feasibility test: Information Book for Telephone Respondents

Mode differences in Interview Survey data were noted above (Section 2.1.3). Safir and Goldenberg (2008) showed that significant differences in expenditure reporting between personal-visit and telephone respondents emerged at the detailed reporting level, and that the use of records and the CE Information Book reduced or eliminated some of these differences. While the Information Book was designed to be shown to respondents during personal visit interviews, the Interviewer Manual contains no guidance concerning its use in telephone interviews. Some Regional Offices direct interviewers to read all the item response categories on the instrument screen to respondents, even though some of the categories include 20 or more items, and even though item categories on the instrument screen do not always match the examples in the Information Book. In addition, an interviewer focus group in 2006 revealed that interviewers are inconsistent in their use of the Information Book during telephone interviews (Edgar, 2006b). Interviewers recognize that if they read all the item categories, the interview will be longer, but if

they read selected examples from the list of items, they may bias the respondent toward reporting those items. Currently available data suggest that most telephone respondents do not use the Information Book.

The CE will shortly conduct a test to determine whether it is feasible to provide a version of the Information Book to respondents who will be interviewed by telephone. An Interagency BLS-Census team produced a lightweight Information Book that can be mailed to respondents when interviewers arrange appointments for telephone interviews in advance. CE currently has no information about how often telephone interviews are pre-scheduled. One objective of this test is to see whether there is enough time to send the Information Book in advance of the interview. Another objective is to see whether respondents will use it. The test will also evaluate the cost of sending the Information Book. Data collection is scheduled to begin in August, 2009 and last for three months. If justified by the test results, CE plans to begin sending an Information Book to telephone respondents in production within the year.

4.3 Bounding

During collection of expenditure data, respondents may make errors in recalling the month and amount of a past purchase (Neter and Waksberg 1964). External *telescoping*, the reporting of purchases made outside the reference period, is one such error. "Bounding" an interview is a technique used to mitigate the effects of telescoping in repeated interviewing. With bounding, information collected in the current interview is compared with information collected in the prior interview. The prior interview serves to truncate reports of events that are remembered as having occurred earlier in time than the reference period of the current interview, and helps to avoid double counting of earlier purchases. Early CE research demonstrated the value of bounding as a way to control overreporting (Silberstein, 1990).

Since valuable resources are expended to collect data on purchases in the first interview, but these data are not used to produce expenditure estimates, a question arises as to how effectively the first interview bounds the second interview. Both the consumer environment and CE data collection procedures and technology have changed considerably since the time of the Silberstein research. This study takes a first step to answering the question by comparing the patterns of reporting between bounded and unbounded second interviews using historical CE data. The study is currently underway.

4.4 Validation/Qualitative Exploratory Study

The goal of this exploratory qualitative study, currently underway, is to understand the Interview Survey process and the challenges faced by respondents at a micro level, in detail that is not practical in a large scale study. Given the large proportion of CE interviews conducted by phone (Safir & Goldenberg, 2008), one element of the study is to look at the possible impact of mode on measurement error.

This study includes a two part interview, and is being conducted by senior researchers in the field. The first part is a CE interview that mimics actual field administration as closely as possible. Participants have been randomly assigned to be interviewed by telephone or in person. Following production procedures, participants receive an advance letter which confirms the date and time of the interview, and provides the same background information about the interview

that is sent to regular Interview Survey respondents. The second part of the interview is an in-depth debriefing interview conducted in the participant's home. There is a 15 minute break between the two parts, during which time participants are asked to collect any records or materials that would help them review their answers to the interview questions.

The debriefing questions are designed to explore several known issues with the Interview Survey, including interview length, recall difficulty, accuracy of responses, record and Information Book usage, telephone interviewing, and proxy reporting. As part of the debriefing, the researchers target specific Interview sections, and ask the respondent to review reported expenditures with any records they might have that would validate (or change) their reports.

Thus far, eight interviews have been conducted and a considerable amount of information has been collected about measurement errors resulting from recall issues, proxy reporting, the differential structure of records and interview questions, respondent expectations and mode. Additionally, with the use of records and follow-up interviews with other household members, a quantification of underreporting has been possible in some instances. The researchers hope to complete an additional ten interviews this summer, and to begin analysis of the data, most of it qualitative, later in the year.

4.5 Order effects

The purpose of this study is to examine the affect of interview length on data quality. Interview Survey interviews average 65 minutes, and respondent and interviewer fatigue can have an impact on data quality. Fatigue is likely to be more of a factor in later sections of the questionnaire, and may lead to satisficing behaviors. The Order Effects test is trying to address the interview length question indirectly by evaluating whether asking a question earlier in the interview will result in higher data quality. The basic idea is to move one section of the questionnaire from a later part of the survey to earlier in the survey, and to analyze the impact of that move on the data estimates. The test will be run using production sample. CE staff are currently making decisions about which section to test, whether to limit the test to the first interview or to include all waves, and the length of the data collection period.

4.6 CE Surveys Redesign: The Gemini Project

The Gemini Project is a CE management initiative with an overall mission of redesigning the CE surveys. The primary redesign goal is to improve data quality, with a particular focus on underreporting. The effort to reduce measurement error will be explored in a manner consistent with combating further declines in response rates, so any expected benefits of survey design changes will be balanced against potential negative effects on response rates.

Gemini was developed in response to the findings presented in the CE Program Review Team Final Report.⁶ The primary objective of the Gemini Project is to develop a detailed research roadmap for a redesign of the Consumer Expenditure (CE) surveys. As presented in the Final Report, the roadmap:

⁶ The CE Program Review was a "top to bottom" evaluation of the CE Program, conducted by senior BLS staff representing all other program areas in BLS. It was conducted between March 2007 and February 2008, resulting in a final report released internally in November, 2008.

will describe the priorities, individual steps, timeframe, resource needs, and costs required for the development, pilot testing, evaluation, and implementation of a redesigned CE survey or surveys. To allow for an unpredictable budgetary environment, the roadmap will address both a complete redesign of the CE surveys as well as more limited modifications to the current design.

A secondary objective of the project is to generate a planned research agenda that will guide the development and implementation of CE research studies throughout the Gemini project lifecycle.

Gemini is proceeding in two stages. The first stage is underway, with development of a management structure that has chartered several teams. Those teams are responsible for:

- Creating a research database that documents all completed, in-progress, planned, and proposed CE research projects (Interview and Diary) which potentially feed into the survey redesign
- Determining what constitutes data quality for CE
- Coordinating a Redesign Conference in 2010 to discuss redesign options
- Assessing the impact of possible redesigns on data user products.
- Presenting redesign models and a detailed project plan for achieving the selected model to BLS management for approval and funding.

A later stage of the project will be to implement the field tests and development projects that result in a redesigned CE.

5. Planned or proposed research

In addition to, or in support of, research deriving from the Gemini Project, the CE program would like to conduct research in a number of areas. Some avenues of future redesign research include:

- **Individual Diaries.** Plans for a pilot test were interrupted in 2008, but the potential gains warrant further exploration.
- **Multi-mode Survey.** Incorporating multiple data collection modes to take advantage of new technologies. Consideration of newer technologies could include forms based Internet collection, Field Representative assisted Internet data collection with avatars or video, UPC scanners, and collecting the output of personal finance software.
- **Alternate Poverty Measures.** Revisions to the Interview Survey to collect additional data on public assistance programs to support the use of the CE as the data source for an alternate poverty threshold. The CE is named in legislation for an alternate poverty measure that has been introduced in the Congress this year.
- **TPOPS.** Testing the feasibility of collecting specific purchase locations in the CE as a way to address increasing response rate concerns about CPI's Telephone Point of Purchase Survey (TPOPS). The TPOPS queries respondents on expenditures, including the name of the outlet where the purchase was made, the type of outlet (on-line, catalog, brick and mortar, etc.), and, the city if it is a brick and mortar location. The CE and CPI are considering a feasibility test of including the outlet locations in the CE surveys for some types of expenses.
- **Combine Interview and Diary samples.** Combining the two surveys into one sample, similar to the project Statistics Canada is undertaking with the Survey of Household Spending. Thus a CU would both be interviewed about their spending, and would be left a diary to record expenses.
- **Reduce overlap between the two samples.** Reducing or eliminating the overlap of expenditure categories between the two CE surveys in order to reduce respondent burden and survey costs. For example, clothing expenditures are a separate section in the Interview survey and also a separate section in the Diary survey. Published estimates for clothing come from both sources—about half from each survey.
- **Targeted Screeners.** Improving the use of group-targeted skip questions—such as identifying households with and without business expenses—to reduce interview length. Groups could also be targeted by income category to skip certain questions about government programs for which they would be ineligible, such as food stamps or public housing.

6. References

- Bates, Nancy. 2006. "Data from the Contact History Instrument (CHI) – 2005 CEQ." Presentation at the 2006 Consumer Expenditure Survey Supervisor's Conference, Tucson, AZ.
- Bradburn, Norman M. and Seymour Sudman. 1991. "The Current Status of Questionnaire Research." In Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman eds.) *Measurement Errors in Surveys*. New York: Wiley, pp. 29-40.
- Creech, Brett, Patrick Falwell, Mary Lynn Schmidt, and Barry Steinberg. 2008. Consumer Expenditure Survey Integrated Table Processing Source Selection Team Final Report. Washington, DC.: U.S. Bureau of Labor Statistics, Internal Report, May.
- Davis, Jeanette, Linda Stinson, and Nhien To. 2002. "What Does it Really Mean to be User-Friendly when Designing an Expenditure Diary?" Alexandria, VA: *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 696-701. <http://www.amstat.org/sections/SRMS/Proceedings/>
- Edgar, Jennifer. 2006a. "Impact of the CE Information Book on Question Comprehension and Recall." U.S. Bureau of Labor Statistics, Office of Survey Methods Research. Internal report.
- Edgar, Jennifer. 2006b. *Field Representative (FR) Focus Group Results: CE Concepts*. Internal Memo submitted September 8, 2008. Washington, DC: U.S. Bureau of Labor Statistics.
- Edgar, Jennifer, Jeanette Davis, Sara Spell, Richard Verlander and Grace Wetzel. 2006. "Individual Diary Feasibility Test. Final Report."
- Goldenberg, Karen L., David McGrath, and Lucilla Tan. 2009. "The Effects of Incentives on the Consumer Expenditure Interview Survey." Internal report. Washington, DC: U.S. Bureau of Labor Statistics, June.
- Gonzalez, Jeffrey M. and Eltinge, John L. 2007a. "Multiple Matrix Sampling: A Review." Alexandria, VA: *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 3069-3075. <http://www.amstat.org/sections/SRMS/Proceedings/>
- Gonzalez, Jeffrey M. and Eltinge, John L. 2007b. "Properties of Alternative Sample Design and Estimation Methods for the Consumer Expenditure Surveys." Presented at the 2007 Research Conference of the Federal Committee on Statistical Methodology, Arlington, VA, November.
- Gonzalez, Jeffrey M. and Eltinge, John L. 2008. "Adaptive Matrix Sampling for the Consumer Expenditure Quarterly Interview Survey." Alexandria, VA: *Proceedings of the American Statistical Association, Section on Survey Research Methods*.
- Gonzalez, Jeffrey M. and Eltinge, John L. 2009. "Imputation Methods for Adaptive Matrix Sampling." To be presented at the Joint Statistical Meetings, Washington, DC, August.

- Groves, Linda S. 2003. "Computer-Assisted Personal Interviewing for the Consumer Expenditure Interview Survey," Consumer Expenditure Survey Anthology 2003. (<http://www.bls.gov/cex/csxanthol03.htm>)
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*. NY: Wiley.
- Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, Robert M. and Steven Heeringa. 2006. "Responsive Design for Household Surveys: Tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society Series A: Statistics in Society*, 169: 439-457 Part 3.
- Henderson, Steven and David Swanson. 2003. "Final Report of the Interview CAPI Test Analysis." Internal report. Washington, DC: U.S. Bureau of Labor Statistics, June.
- Henly, Megan and Nancy Bates. 2006. "Using Call Records to Understand Response in Panel Surveys." Alexandria, VA: Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 4128-4135. <http://www.amstat.org/sections/SRMS/Proceedings/>
- McGrath, David E. 2005. Comparison of Data Obtained by Telephone versus Personal Visit Response in the U.S. Consumer Expenditures Survey. Alexandria, VA: Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 3368-3375. <http://www.amstat.org/sections/SRMS/Proceedings/>
- McGrath, David, Boriana Chopova, Christie Gomes, and Nathan McDermott. 2007. "The Effects of Incentives on the Consumer Expenditure Diary Survey. Final Report." Washington, DC: U.S. Bureau of Labor Statistics, August.
- McGrath, David E. and Lucilla Tan. 2007. "Does the time-of-day or day-of week when interviewers contact respondents affect respondent cooperation? Findings from the Consumer Expenditure Interview Survey." Draft internal report. Washington, DC: U.S. Bureau of Labor Statistics, December.
- Neter, John and Joseph Waksberg. 1964. "A Study of Response Errors in Expenditures Data From Household Interviews." *Journal of the American Statistical Association* 59.
- Office of Management and Budget. 2006. *Standards and Guidelines for Statistical Surveys*. Washington, DC: U.S. Office of Management and Budget.
- Ortutay, Barbara. 2009. "Survey: Family time eroding as Internet use soars." Associated Press Report in *The Washington Post*, June 15.
- Paulin, Geoffrey, Jonathan Fisher, and Sally Reyes-Morales. 2006. *User's Guide to Income Imputation in the CE*. U.S. Bureau of Labor Statistics, Division of Consumer Expenditure Surveys.

Piani, Andrea. 2004. "Contact History Initiative. Washington, DC. Presentation at the 2004 Federal CASIC Workshop, Session on the Use of Field Metrics and Paradata (Process Data) in Evaluating CASIC Surveys.

Raghunathan, T.E. and Grizzle, J.E. 1995. "A Split Questionnaire Survey Design." *Journal of the American Statistical Association*, 90:54-63.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Safir, Adam and Goldenberg, Karen. 2008. "Mode Effects in a Survey of Consumer Expenditures. Alexandria, VA: Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 4436-4343. <http://www.amstat.org/sections/SRMS/Proceedings/>

Silberstein, Adriana R. 1990. "First wave effects in the U.S. Consumer Expenditure Interview Survey." *Survey Methodology* 16: 293-304.

Taylor, Paul, Richard Morin, D'Vera Chon, Richard Fry, Rakesh Kochhar, and April Clark. 2008. *Inside the Middle Class: Bad Times Hit the Good Life*. Washington, DC: Pew Research Center, Social and Demographic Trends Report. <http://pewsocialtrends.org/assets/pdf/MC-Executive-Summary-and-Overview.pdf>

To, Nhien and Lucilla Tan. 2008. "Evaluation of the 2005 Redesigned Consumer Expenditure Survey Diary." *Consumer Expenditure Survey Anthology 2008*. <http://www.bls.gov/cex/csxanthol08.htm>

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Tucker, Clyde. 1985. "Identifying Measurement Error in Consumer Expenditure Reports." Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 476-481. <http://www.amstat.org/sections/SRMS/Proceedings/>

Tucker, Clyde, Brian Meekins, and Paul Biemer. 2006. "Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Further Micro-Level Latent Class Analysis." Washington, DC.: U.S. Bureau of Labor Statistics, Internal Report, August.

Appendix A: Revisions to the CE Interview Survey, 2004 through 2009 and Proposed for 2011

2004 Changes:

- Deleted the quantity consumed and the unit of measure questions for utilities
- Added a separate question for money put into educational savings accounts (e.g. 529 plans)
- Expanded the question about “haircutting” to include manicures and other salon services

2005 Changes:

- Added additional help screens to the CAPI instrument
- Consolidated and modernized the list of appliances inventoried and collected
- Modernized the types of “technology” items and services collected such as televisions, DVD players, personal music player such as Ipods, PDAs, and fees for services that use Global Positioning Service (GPS) such as OnStar
- Streamlined questions for collecting home equity loan payments
- Added references to online and automated payments as cues to respondents to help them remember purchases made this way, particularly for services
- Created a new section to collect internet services such as downloading audio or video files and online gaming
- Created a better mechanism for collecting combined clothing expenditures, specifically, allowed the interviewer to collect information about which persons the combined purchase of clothing was for
- Deleted questions that were no longer needed by data users in the vehicles sections
- Included a reference to reimbursements from Flexible Spending Accounts in the medical reimbursements section
- Improved the structure for collecting subscription, membership, and entertainment expenditures
- Introduced new informational flyers for both surveys

2006 Changes

- Simplified the collection of telephone services by eliminating many of the detailed check boxes no longer needed by users
- Deleted questions no longer needed by users in the vehicle sections
- Created a new section to collect data on CU member enrollment in the Medicare Prescription Drug plan, Medicare Part D

2007 Changes

- Re-consolidated and re-modernized the list of appliances inventoried and collected
- Improved the questions related to rent as a form of pay
- Added a question on whether the rental unit is under a rent control program

- Improved the collection of data on owned timeshare, vacation, and multi-owner properties (added more detailed question about percentage of ownership, location of property, and specific prompts for timeshares)
- Improved the collection of data on mortgages, including the collection of interest only and variable rate mortgages
- Improved collection of rental equivalence for timeshares and second homes or vacation properties
- Started collection of the property value of owned properties every quarter instead of only once
- Added a specific code for the collection of Voice over Internet Protocol (VOIP) services in the telephone section
- Completely redesigned the structures of the telephone and utilities sections to reduce the amount of time to collect the data in these sections, reduced the likelihood of “double reporting” data from quarter to quarter, improved the collection of combined services items, and eliminated questions no longer needed by users
- Added a specific question for the collection of prepaid cellular phone services
- Added a specific item code for Satellite radio services
- Eliminated the collection of names in the medical section to protect respondent privacy
- Included a reference to reimbursements from Medical Saving Accounts in the medical reimbursements section
- Modernized the collection video and computer game hardware and software
- Improved the reference period for the collection groceries and meals away from home
- Added questions about reverse mortgages
- Improved question wording throughout the survey

2008 Changes

- Added questions to capture the receipt, amount, and primary use of the 2008 economic stimulus payments

2009 Changes

- Discontinued asking for a CU member's day of birth to ask only for the month and year of birth to protect respondent privacy.
- Added an introduction to the survey and to the clothing section (Section 9A) reminding the respondent to refer to bills during the interview.
- Clarified or simplified the wording of many questions throughout the survey.
- Created a separate item code for expenditures on "portable media."
- Moved the item codes for "Care for invalids, convalescents, handicapped, or elderly persons in the home" and for "Adult day care centers" from Section 19 to Section 15 so we can also capture reimbursements for these items.
- Discontinued asking for member names in the education expenses section to protect respondent privacy.
- Created a separate item code for "test preparation or tutoring services."
- Added questions to capture the receipt and use of the one-time \$250 stimulus payment for Social Security recipients and other retirees

Proposed for 2011

- Improve the collection of rental equivalence by improving the current questions related to timeshares and vacation homes available for rent and better identifying rental units contained within a sample unit such as a basement apartment
- Add new questions to support alternative poverty thresholds, such as more detail on school meals, subsidized utilities, and the Earned Income Tax Credit
- Improve collection of bundled technology services (e.g. telephone, internet, and television) and equipment such as phones and satellite receivers and modems
- Reduce respondent burden when reporting health insurance premiums by calculating, rather than re-asking, premium amounts
- Streamline the household repairs and trip sections
- Introduce new products and services

Appendix B: CE Source Selection

CE Source Selection is a multi-step process for determining which source—Interview Survey or Diary Survey—to use for the published integrated tables. The process uses three years of data and gives more weight to the most recent year's data. Source Selection for 2007 used 2004 through 2006 data. The counts and z-scores are weighted for the three most recent collection years using the following breakdown, which places greater emphasis on the more recent collection years. The weight for the earliest collection year is one-sixth, the second collection year is two-sixths, and the most recent collection year is three-sixths. However, if a new expenditure category code was created within the most recent two years, then the two most recent years of data are used, and the weight for the first collection year is two-fifths and the weight for the most recent collection year is three-fifths.

After weighting the data, the next stage focuses on counts. One or both of the surveys may fail to capture a large enough number of expenditures to conclude which survey provides better results. These counts are used to ensure that a sufficient amount of data is available to make sound source selection decisions. A “sufficient” amount of data exists when the lowest count for any of the three years is greater than or equal to 60. If both surveys have sufficient data, then proceed to the next stage involving z-scores. If both surveys lack sufficient data, then keep the incumbent source.

The next stage determines which survey provides the greater overall expenditure per expenditure category. Compare the means of reported expenditures, weighted by year, from each survey using a standard z-test. If the z-score of a given expenditure category falls outside the range of -1.645 to 1.645, select the survey with the higher mean expenditure due to the statistically significant difference in means. If the z-score of a given expenditure category fell below -1.000 and 1.000, select the incumbent survey as the source due to the statistically insignificant difference in means. If the z-score of a given expenditure category falls between -1.645 and -1.000 or 1.000 and 1.645, an additional criterion for source selection is necessary. Review the z-scores for all of the three years in question. An incumbent source changes only if the z-scores for each of the past three years were greater than 1.000 (use the Interview Survey) or less than -1.000 (use the Diary Survey) (Creech et al, 2008).