Redesign Options for the Consumer Expenditure Survey

November 10, 2011

Prepared for:
The National Academy of Sciences
500 Fifth Street NW
Washington, DC 20001

Prepared by:
Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Westat®

Authors

The principal authors of this report are David Cantor, Sid Schneider, and Brad Edwards. Bob Fay and David Judkins performed the analysis of the impact of the proposed Quarterly Interview survey on sampling variances. Lou Rizzo assisted in the computation of statistical power of the field tests. Pat Cunningham and Pat Ward assisted with developing the cost estimates.

Table of Contents

Appendixes

Executive Summary Tables

Tables

Tables (continued) **Page**

Tables (continued)

Appendix Tables

Executive Summary Figure

Figures

Figures (continued) **Page**

Executive Summary

Introduction

The Consumer Expenditure Survey (CE) program, consisting of the Diary survey and the Quarterly Interview survey, is an ongoing source of data about consumer purchasing in the United States. The program also provides data crucial to the calculation of the Consumer Price Index.

Estimates of consumer expenditures derived from CE data have tended to be lower than estimates derived from other sources. This discrepancy has grown larger in recent years. In addition, the response rate to the survey has declined significantly over the previous 10 years. These issues potentially increase the error in the estimates, as well as reduce the efficiency of the survey. Therefore, the Bureau of Labor Statistics (BLS) recently contracted with the National Academy of Sciences (NAS) to recommend approaches to the redesign of the CE program. NAS contracted with Westat to make recommendations for this redesign, with the goals of reducing measurement error; aligning the program with developments in the retail industry; and offering the respondents more flexibility in the way that they gather and report their data. In this paper, we propose several approaches to redesigning the CE program that address each of these goals.

General Approach of Redesign

The recommended redesign addresses the NAS objectives through three interrelated changes. The first recommended change is to reduce respondent burden. The burden on CE respondents is considerable, which may reduce the quality of the information collected, and make some respondents unwilling to cooperate. For example, the Quarterly Interview survey averages approximately one hour, is repeated five times over a 12-month period, involves asking respondents to recall very detailed information for extended periods of time, and requires a single respondent to proxy for all other household members. All of these requirements lead to significant time and cognitive burden.

A second recommended change to the design is to incorporate administrative and personal record information into the data collection, with the goals of improving data quality and reducing respondent burden. The third recommended change is to improve the methods used to collect the self-report information. Since the CE was designed in the early mid-1960s, there has been significant

progress in designing surveys. This includes reducing the amount of proxy reporting, using various types of incentives and improving the design of the recall interview. Also, many new technologies for collecting data have emerged in recent years.

This report is divided into three parts. The first part reviews literature on the potential role of alternative data sources. The second part provides the specific recommendations for a redesigned CE program. The third part sets forth potential evaluation studies to assess the impact of the proposed redesigned CE program.

Alternative Sources of Data

The proposed redesign of the CE program calls for less reliance on the respondents' memories of their expenditures and greater reliance on records. Records of the respondents' expenditures are potentially available from several sources.

Other data collection activities. First, data aggregators such as Neilsen and NPD obtain price and sales data from a large number of retailers. Neilsen claims to collect data from "virtually every major retail chain." However, data aggregators collect little data from small retailers. Moreover, data aggregators provide little information about the statistical properties of the data. Data aggregators may therefore not be a suitable source of data for the CE program.

Instead, the CE program might collect sales and price data directly from a sample of retailers and estimate expenditures for the purposes of the CPI. The retailers would have to provide microdata tailored to the needs of the CE program. Retailers might seriously consider providing such data if they were asked by the BLS.

The National Consumer Panel is a joint venture of Nielsen and SymphonyIRI involving over 52,500 households. Respondents in the households scan the bar codes of the products that they purchase. Although the National Consumer Panel is intended to collect comprehensive data about consumer expenditures, its methods are in many ways incompatible with the CE program. For example, documentation is absent on sampling, self-selection, nonresponse, and attrition. No studies have been conducted to measure the effect of any bias. The Department of Agriculture rejected the National Consumer Panel as a source of data about food purchases. It is probably also not suitable for the CE program.

The Department of Agriculture elected instead to recruit its own panel, which will provide data about food obtained for consumption at home or away from home. The methods of this program, known as the Food Acquisition and Purchase Survey (FoodAPS), appear to be compatible with the goals of the CE program in many ways. In some ways, however, the methods are not well suited to the CE program; for example, the data collection period and the sample size are not a good fit. Discussions between BLS and the Department of Agriculture seem to be warranted to determine whether the methods of FoodAPS could be adjusted to serve the purposes of both agencies.

Data mining firms such as Teradata and Catalina Marketing work with loyalty card data for great numbers of consumers. Catalina Marketing, for example, claims to work with data from 300 million retail transactions per week. These data mining firms do not analyze the spending of individual households. The potential utility of the data that data mining firms could provide to the CE program is unclear. However, BLS might be able to collect data about respondents' households directly from retailers that have loyalty card programs.

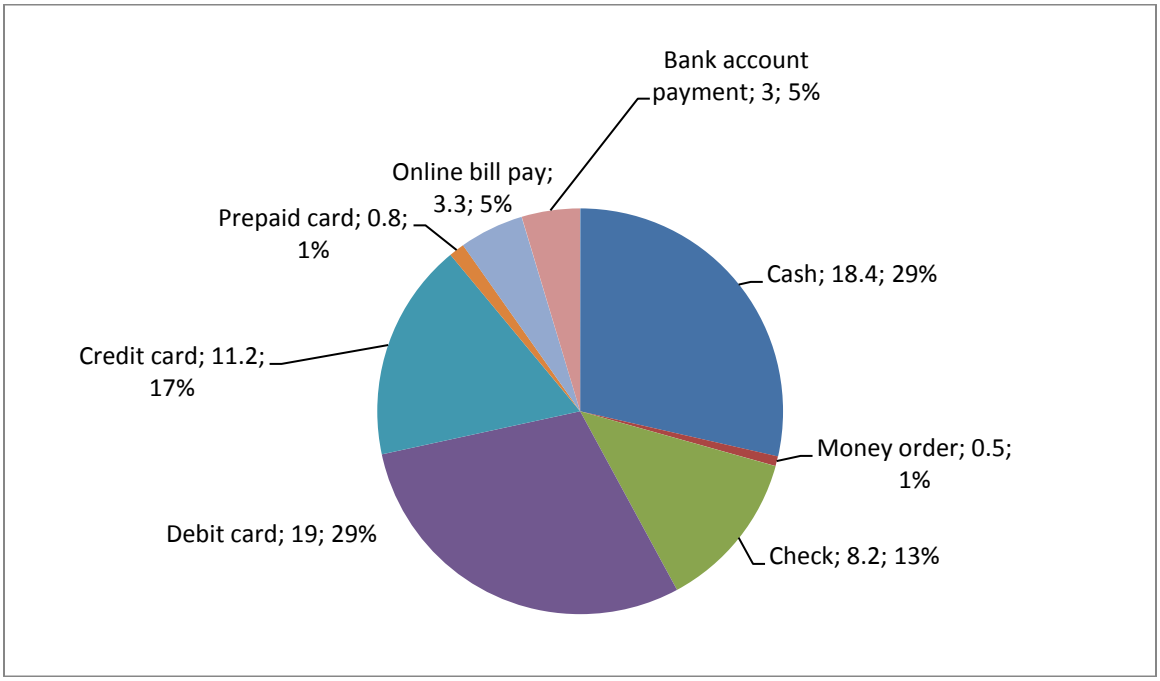
Administrative Data. Data obtained directly from retailers are likely to be more accurate than respondent-provided data are likely to be. Other federal surveys, such as the National Immunization Survey and the Residential Energy Consumption Survey, have employed administrative data to supplement and improve the quality of data reported by respondents.

Conceivably, CE respondents could provide their loyalty card numbers to interviewers, who would then ask the retailers to provide the purchasing histories for those loyalty cards. This method would not be perfect; a consumer may sometimes forget to give a loyalty card to the cashier or may lend the card to friends. Moreover, retailers do not routinely release purchasing histories. The BLS might explore the feasibility of obtaining purchasing history data by contacting large retailers with loyalty card programs. Expenditure data is also potentially available from utility companies, rental agents, and lenders.

Respondent-supplied administrative data. CE respondents would need to expend little effort to access electronic records for their purchases made by check, electronic fund transfer, credit card, debit card, and online payment service. They could download the files directly from the web sites of financial institutions, or use financial software packages such as Mint. Alternatively, respondents could use software that was developed specifically for the CE program. These data files would contain a portion of the data required by the CE program, such as the total amount and the approximate date of purchases.

Data files would usually not exist for purchases made by cash or by money order. However, the results of a survey by the Federal Reserve Bank of Boston revealed that only 29.3 percent of purchases are by cash or money order. The remaining purchases—over 70 percent—are transacted by credit card, check, debit card, or other method that creates an electronic record.

Figure A. Transactions per consumer per month



Respondents therefore can download files that potentially document over two-thirds of their purchases. The data would not be complete, however. The respondents would have to provide receipts or other documentation, or rely on their memories, to fill in the data required for the CE program.

Records of expenditures. Research from the Medical Expenditure Panel Survey, the National Immunization Survey, and other federal data collection efforts suggest that the quality of survey data is enhanced when respondents provide administrative data. In the case of the CE program, respondents could provide receipts that list the description and price of each individual product or service purchased, and the date of the purchase. Respondents who had no receipt for a purchase could create documentation by completing a short form. Alternatively, they could create documentation using a smart phone, by completing an on-screen form, photographing the purchase, or recording a voice memorandum at the time of the purchase.

Recommended Redesign: Diary Survey

We recommend that the CE program continue to collect data in separate Diary and Quarterly Interview surveys. Data collection in our proposed redesign of the Diary survey would follow this schedule:

Table A. Schedule for the redesign of Diary survey

Initial interviewer visit to the household – In person ■ Respondent signs consent form to obtain information electronically ■ The method that respondent will use to provide data is chosen (paper or electronic) ■ Interviewer provides a scanner to respondents who will report electronically ■ Interviewer provides an envelope or box for respondents to store receipts ■ Interviewer provides diary forms to respondents who will neither mail nor scan receipts
2-3 days after initial visit – Telephone ■ Interviewer calls respondent, helps respondent overcome any barriers to providing data ■ Interviewer continues to monitor respondent's data reporting activity
7 days after initial visit – Telephone/In person ■ If the respondent appears to be sending data electronically or on paper diligently, interviewer telephones; otherwise interviewer visits in person ■ If interviewers visit in person, administer a recall interview ■ Interviewer continues to monitor respondent's data reporting activity
10 days after the initial visit – Telephone ■ Interviewer calls respondent, helps respondent overcome any barriers to providing data ■ Interviewer continues to monitor respondent's data reporting activity
14 days after the initial visit – Telephone/In person ■ If the respondent appears to be sending data electronically or on paper diligently, interviewer telephones; otherwise interviewer visits in-person ■ Interviewers who call close out the data collection ■ Interviewers who visit in person administer a recall interview and close out the data collection

The interviewer first visits the respondent, explains the survey, and helps the respondent choose a data collection technique. Every adult in the household is asked to provide expenditure data. The respondent provides the household's loyalty card numbers and signs a consent form for the interviewer to collect purchasing history data from the retailers. The interviewer supplies a box or envelope for the respondents to store receipts and other purchase documentation.

There are two alternative methods that respondents could use to report their data. They could report their expenditure data electronically or on paper. If they provide data electronically, they are given a small scanner. They then scan their receipts and send the scan files to a central repository. They also download data files for their checking, credit and debit cards, and other accounts from their financial

institutions' web sites, and send these data files to the central repository. The repository system would extract purchase data from the receipts using optical character reading technology and software specially designed to extract data about individual purchases. The repository system would then generate a web survey that would ask the respondents to provide the remaining information that the CE program needs.

For example, if a respondent purchased 10 items at Macy's and paid with a credit card, the respondent would send a scan of that receipt and the credit card data file to the central repository. The repository system would match the entry for the purchase in the credit card data file with the receipt. The web survey would include questions such as "On July 5, 2011, you purchased shoes for $55.25 and paid by credit card. Was this purchase for a male or a female?"

Respondents who do not provide data electronically would mail the envelopes containing their receipts to the central repository. They could also mail photocopies of monthly statements for their credit card, checking, and other accounts. Staff at the repository would scan the receipts and enter data from the statements. The system would generate a telephone survey for the interviewers to collect the remaining data needed by the CE program.

The interviewers also collect purchasing histories from retail chains, using the loyalty card numbers for the respondents' households. Of course, retail chains would have to be willing to provide these data.

The interviewers monitor the activity of each household. The interviewers telephone the households and encourage them to provide data, whether electronically or by U.S. mail. They determine whether any respondents are having difficulty providing data and attempt to resolve any problems. These calls occur about 2 days after the initial visit, and again 7 days after the initial visit.

If a household was not providing data adequately, an interviewer would visit the household in person 7 days after the initial interview. The interviewer would conduct a recall interview, in which the respondent would report the household expenditures during the 1-week reporting period. The interviewer would use an Event History Calendar technique to facilitate the respondent's recall.

Ten days after the initial interview, the interviewers again telephone the respondents. The interviewers again try to resolve any difficulties that the respondents might be experiencing and encourage the respondents to provide their data.

Fourteen days after the initial interview, the interviewers close out the data collection. If a household appears to have provided adequate data, electronically or by U.S. mail, this final contact is by telephone. Otherwise, the interviewer visits the household and conducts a recall interview. Again, the interviewer uses an Event History Calendar technique.

Using data on the cost of the current Diary survey, we estimate the recommended design will cost approximately 60 percent more than the current design ($8 million versus $5 million). Most of this increase is attributable to the use of multiple respondents in each household. A smaller portion is from the administrative record component.

Recommended Design: Quarterly Interview Survey

Data collection for our proposed redesigned Quarterly Interview survey would follow this schedule:

Table B. Quarterly interview survey schedule and procedures for sample CUs

Initial interview - In-person interview
- Obtain consents
- Determine data collection methods, electronic or paper
- Provide data collection instructions
- Conduct bounding interview

2 weeks, Month 1 and Month 2 after initial interview – Telephone c ontact
- Review respondent's provision of data, electronic or paper
- Encourage respondent to provide data

3 months after initial interview – In-person contact
- Review records collected
- Conduct recall Interview

12 months after initial interview – Telephone/In-person
- Mail instruction packet beforehand
- Followup telephone call to orient respondent and conduct bounding interview
- If household has changed, use initial contact protocol

12 months plus 2 weeks, 13 months, 14 months after initial interview – Telephone contact
- Review respondent's provision of data, electronic or paper
- Encourage respondent to provide data

15 months after initial interview – In-person interview
- Review records collected
- Conduct recall Interview

We recommend only two waves of data collection, instead of the current four. Each wave would last 3 months. Data about food and alchohol expenditures would be collected in summary form. The Diary survey would be used to estimate these expenditures for the CPI.

As in the Diary survey, respondents would be asked to furnish their loyalty card numbers so that administrative data could be collected from retailers. In addition, utility and mortgage companies would be asked to provide expenditure data on items such as electricity, water, gas and housing. Unlike the Diary survey, only one respondent would provide data per household.

As in the Diary survey, respondents could provide their data electronically. They could scan their receipts and send the receipt files to a central repository. They could also send the data files that they download for their credit card, checking, and other accounts, or they could send their receipts and photocopies of their monthly account statements by U.S. mail. The repository system would automatically extract the purchase data and generate a web survey or a telephone survey script.

The interviewer would telephone the respondents 2 weeks, 1 month, and 2 months after the initial interview to encourage the respondents to send in their data, and to resolve any difficulties that the respondents may have encountered.

Three months after the initial interview, the interviewer would conduct an in-person interview. This contact would include a recall interview if the respondent had not provided adequate data. Again, the recall interview would employ an Event History Calendar technique. The reference period need not be uniformly 3 months. We recommend that the reference period be 1 month, 3 months, or 12 months, depending on the category of the expenditures. We recommend that the BLS continue to conduct research to identify the optimum recall periods.

One year after the initial contact, the process is repeated for the second wave. The initial contact for the second wave is by telephone, unless the household had changed. In that case, the contact is in person.

Our recommended redesign for the Quarterly Interview survey would cost nearly twice as much as the existing methods. The increase is attributable to the additional effort in contacting more households. Reducing the number of panel waves from four to two requires twice as many households to collect the same amount of information. However, ththis design also results in a significant reduction in sample variation, due to eliminating the within-household correlation across panel waves. For example, maintaining the current 3-month recall period for all expenditures results

in a median reduction in sample variation of 35 percent over the current design. This variation can be further diminished by selectively increasing the reference period to 12 months for appropriate items.

The other major cost increase is attributable to the administrative survey of retailers and utility companies. This accounts for approximately half of the increase. It is possible to significantly reduce the cost of this survey by restricting the survey to utility companies. If only utility companies are involved, the cost would be about 30 percent of the cost of including retailers as well.

In summary, our recommended redesigned Diary and Quarterly Interview surveys compare to the existing methods as follows:

Table. C. Comparison of design features of current and redesigned surveys

Design feature	Current collection	Redesigned survey
Diary Survey		
# of Recall Interviews/CU	0-2 (depends on use of diary)	0–2 (depends on record use)
Reference period	1 week	1 week
Sample design	Cross-sectional	Cross-sectional
Mode	Paper	Web and paper
In-person/telephone interviews	3	1 to 3, depending on use of records
Telephone reminders	Yes, unknown how many	1–4 depending on use of records
Number of respondents	1 person	All persons 14+
Incentives	None	Monetary and nonmonetary
Personal electronic records	None	All respondents willing to use it
Use of receipts	Yes, not strong emphasis	Yes, very strong emphasis
Administrative Record Survey	No	From retailers used by respondents
Interview Survey		
# of Recall Interviews/CU	4	2
Reference period	3 months	1, 3, 12 months depending on item
Sample design	Panel – 4 waves (3 months apart)	Panel, 2 waves (12 months apart)
Mode	CAPI	Web, CAPI
In-person/telephone interviews	5	3
Telephone reminders	Not known	6–7, depending on use of records
Number of respondents	Main CU respondent	Main CU respondent
Incentives	None	Monetary and Nonmonetary
Personal electronic records	None	All respondents willing to use it
Use of receipts	Yes, not strong emphasis	Yes, very strong emphasis
Administrative Record Survey	No	From retailers and utility companies

Our recommended redesign would meet the needs of CE analysts, as described below.

Table D. Data requirements for CE analysts and redesigned CE program

Requirement	Addressed by redesign?
Complete picture of spending	Yes. If modules are used on interview, imputation might be required
Level of expenditure detail	Yes
Population	Yes
Sample size and geographic detail	May need to increase cost to meet current requirements
Time period and periodicity	Yes
Panel component	Annual estimates of change. It will not be possible to estimate at a CU level: (1) expenditures for an entire year; (2) change for intervals of less than 1 year.
Integrated microdata	Yes
Income	Yes, but not directly addressed in recommendations
Socioeconomic information	Yes
Assets and liabilities	Yes, but not directly addressed in recommendations

We recognize that BLS is not likely to be in a position to spend twice as much on the survey in the near future. In light of this, we suggest priorities for the components of our proposed redesign. The highest priority should be given to the personal record component, which offers the greatest promise to improve measurement and reduce respondent burden. It is also the least expensive component associated with our recommendations. We also recommend that the use of both the multiple diary keepers and the incentives be tested. If they are found to reduce total survey error, they should be adopted. Both have the potential to improve measurement and reduce burden. We recommend the next highest priority should be to conduct some type of administrative survey, giving a survey of utility companies the highest priority. This should be the least expensive to administer and offers a direct substitute for a substantial portion of the current CE interview. A design that reduced the number of panel waves is seen as very desirable from the point of view of reducing respondent burden. This may be affordable if future research shows that it is possible to incorporate a 12-month reference period for selected items that are now collected with a 3-month period. This would result in a significant reduction in sampling variance and make our proposed design more economical.

Development and Evaluation of the Proposed Redesign

We propose methods for developing the software and other materials needed for our proposed redesigned CE program. We propose a series of small field tests to evaluate the software, such as the software to scan the receipts and send the scan files to the central repository, and to extract the purchase data and generate the followup surveys. The field tests would also evaluate printed materials, such as instructions and forms for recording purchases that have no receipts. Field tests could also evaluate methods for assigning data collection techniques to individual respondents, and for collecting administrative data from retail chains, utilities, rental agents, and others.

A field test of the proposed redesigned Diary survey could use this design:

Table E. Experimental design for field test for the Diary survey

Response method	New design – respondent selection				Current design
	Single respondent		Multiple respondent		
	$50	$70	$50/$10	$80/$20	9
Method 1	1	2	3	4	
Method 2	5	6	7	8	

Households would be assigned to have one or multiple respondents; to have a low or high value incentive; and to either have a data collection method assigned or to allow the respondents to choose a data collection method. These conditions would be compared with the current design. Outcome variables would include the mean level of expenditures, the proportion of expenditures that match retailer administrative data, the proportion of expenditures that respondents report using records rather than memory, nonresponse, cost, time required for data collection, and proportion of expenditures reported during the recall interview.

A field test of the proposed redesigned Quarterly Interview survey could use this design:

Table F. Design for field test for interview survey that includes reference period as experimental factor

Reference Period Group	New design – respondent selection				Current design
	Single respondent		Multiple respondent		
	$50	$70	$50/$10	$80/$20	9
Group 1	1	2	3	4	
Group 2	5	6	7	8	

Households would be assigned to have one or multiple respondents; to have a low or high value incentive; and to have fixed reference periods or reference periods that vary depending on the expenditure category. These conditions would be compared with the current design.

Our report discusses the number of households to include in the field tests and provides a power analysis.

Introduction 1

The Consumer Expenditure Survey (CE) program is an ongoing source of data about consumer purchasing in the United States. CE data have a crucial role in the calculations for the Consumer Price Indices (CPI) and other statistics. The CE program also provides extensive data about the characteristics of consumers themselves. Researchers use CE data to monitor consumer behavior and to study issues such as the relationship between consumers' demographic backgrounds and spending.

The design of the CE program must be periodically updated to take advantage of new technologies and to ensure that CE methods are aligned with changes in the retail environment and survey practice. Recently, the Bureau of Labor Statistics (BLS) contracted with the National Academy of Sciences (NAS) to review the redesign options available and to identify the most promising ones.

The redesign is motivated by several issues. The first and most important issue is measurement error. Comparisons with other sources of data suggest that CE respondents tend to underreport their household expenditures. This underreporting may be attributable to several causes. For example, CE respondents must report for everyone in the household. Respondents may find this task to be overly burdensome and may not thoroughly gather or report expenditure data from others in the household. Other potential sources of error in the CE program are panel conditioning, frequent telephone administration of surveys that were intended to be administered face-to-face, and the length of the reporting period.

The second issue pertains to changes in the survey and retail environments. Consumers are increasingly making purchases online or through automatic fund transfers. Also, they often buy many different types of merchandise in a single purchase at "big box stores" such as Target. The current CE methods may not adequately accommodate these expenditures. Moreover, new technologies may offer innovative methods for collecting expenditure data. For example, smartphones may provide a way for some respondents to track their purchases; financial software packages provide a means for respondents to download their credit card and checking account data.

The third issue is the need for greater flexibility to reflect the diverse preferences of the respondents. Current CE methods do not sufficiently tailor data collection methods to the respondents. A multi-

method, multi-mode approach might allow respondents to use the data collection method that is best suited to their preferences, their daily routines and their abilities.

This report presents a number of approaches that BLS might pursue to redesign the CE program in a way that addresses these issues. The goals of the report are to analyze the challenges that the CE program currently faces, and to propose solutions that can improve the ability of the CE program to measure expenditures by U.S. consumers. The report also proposes an evaluation study to assess the feasibility of the proposed solutions.

This report is organized into chapters:

- Chapter 2 discusses the objectives of the redesign of the CE program: to reduce measurement error, to take into account changes in the retail industry and the development of new electronic tools, and to provide respondents with greater flexibility;
- Chapter 3 presents potential approaches to redesigning the CE program, with an emphasis on reducing reliance on the accuracy of the respondents' recollections about their expenditures;
- Chapters 4 and 5 present the proposed redesign of the Diary survey and the Quarterly Interview survey, respectively, including discussions of the implications of our recommendations for cost and precision; and
- Chapter 6 presents several potential approaches to evaluating our redesign recommendations.

2 Objectives of the Redesign

Comparisons between CE data and data from the Personal Consumption Expenditures (PCE) of the National Income and Product Accounts, produced by the Bureau of Economic Analysis, suggest that CE respondents tend to underreport expenditures. Estimates of consumer expenditures from the CE program are generally lower than estimates from the PCE. Some of this discrepancy is probably due to methodological differences between the two data collection programs, as Garner et al. (2006) discussed:

> Simply put, the Bureau of Labor Statistics collects CE expenditure data through sample surveys and weights the results to obtain population estimates. The Bureau of Economic Analysis, in contrast, calculates PCE estimates on the basis of industry production data collected in economic censuses and through surveys conducted by outside agencies. There are clear differences in the types of expenditure data obtained, dictated by the data collection methods and data sources used by the two agencies. In addition, the populations covered by the CE and the PCE differ.

Even considering these methodological differences, comparisons between the CE and PCE data strongly suggest that CE respondents tend not to report their expenditures completely. The CE-to-PCE ratio for comparable expenditures is approximately 80 percent overall. This ratio ranges between 60 and 70 percent for durable and non-durable goods, while it is over 90 percent for services (Garner, 2009). These ratios have been declining over time. For example, for comparable items of durable goods, the CE-to-PCE ratio declined from 88 percent to 69 percent between 1992 and 2007 (Garner et al., 2009).

Sources of Measurement Error

Some of this decline is very likely due to increasing error in the CE. The response rate for the CE interview survey has declined steadily from 82 percent in 2000 to 74 percent in 2010 (Safir, 2011). This trend tracks closely the general decline in response rates for other general population surveys (Brick and Williams, 2009).

In addition, an increasing number of CE interviews are being conducted by telephone, instead of face-to-face. Respondents tend to report fewer expenditures when they are interviewed by telephone (Safir and Goldenberg, 2008). In a face-to-face interview, the interviewers employ show cards to help the respondent recall expenditure items; they cannot use show cards during a telephone interview. Past research suggests that telephone respondents are more likely than face-to-face respondents to "satisfice," or issue a response that is not very accurate but which moves the survey along (Hollbrook et al., 2003). "Satisficing" is especially likely with challenging interviews that require a great deal of recall, like the CE interview.

Respondent burden may be an additional source of measurement error in the CE program (Lessler, 1987; Cantor, 2010). First, respondents may find the task of recalling their expenditures to be inherently burdensome. While in theory the Diary survey respondents are expected to log their expenditures as they occur, in practice many respondents recall expenses as best they can immediately before the interviewer collects the diary.

The 3-month reference period in the CE Quarterly Interview survey may also be a source of respondent burden and measurement error. A classic study by Neter and Waksberg (1964) found that the length of the reporting period had a significant effect on the accuracy of the expenditure data that respondents provided. Shorter, bounded, reference periods produced significantly more reports of expenditures than longer, unbounded periods. The extent of this effect varied depending on the nature of the expenditure. Respondents may not remember small and unremarkable expenditures over a 3-month period. Respondents may have particular difficulty recalling the required supplementary information, such as the price and the taxes paid. For example, in 2007 the CE-to-PCE ratio was 96 percent for automobiles, but only 44 percent for video and audio goods (Garner et al., 2009). One possible reason for this difference is that the purchase of an automobile is a more salient and memorable event.

The recall task is made more complex by the requirement that CE respondents report on all expenses for all members of the consumer unit (CU). While respondents are encouraged to consult other household members during the reference period, household members may not be motivated to give the respondent accurate data consistently. The respondents can report only on the expenditures that they know about, and their knowledge may be incomplete (Kojetin and Miller, 1993). Also, the respondent may be reluctant to report on certain expenditures in the consumer unit, fearing that reporting the expenditure would be a violation of the privacy of another member of the household.

The sheer number of questions in the Quarterly Interview survey could add to respondent burden. The interview is quite long, and likely to fatigue some of the respondents, who may answer less carefully as the interview goes on. The published length of the interview survey is one hour. Of course, the actual length would be expected to vary according to the number of expenses to be reported. Larger and more affluent households may have more expenses, so the respondents may need more time to complete the interview. Because each respondent is interviewed five times over the course of one year, the real and perceived burden is great.

The term "panel conditioning" refers to the situation in which a respondent deliberately answers a question in a way that avoids additional questions. For example, the respondent might untruthfully say that the household had no expenditures of a particular kind, to avoid having to answer questions about those expenditures. With long interviews such as the CE Quarterly Interview, panel conditioning is a possible contributor to measurement error.

Interviewer burden is an often overlooked phenomenon which could also have a role in the quality of CE data. If CE respondents appear to have difficulty providing data, to take shortcuts, and to be annoyed, the interviewer may take steps to shorten the interview. CE interviewers conceivably may be inclined to accept partial answers or to encourage respondents to work through the interview quickly at the expense of data quality.

Any redesign of the CE surveys must take into account the need to reduce respondent burden and measurement error. The redesign must balance the need for a great amount of detailed data with the need to ensure that the respondents report those data completely and accurately.

Electronic Records of Expenditures

Consumers today commonly make purchases using modes that leave an electronic record. When an electronic record exists, respondents potentially could provide the expenditure data by retrieving information about the purchase from a database, or by printing out a record of the purchase, rather than by trying to remember the details of the purchase or by finding a receipt.

For example, soon after a consumer makes a purchase using credit card, debit card, check, electronic fund transfer, or PayPal, a record of the transaction appears in a file that can be downloaded from the web site of a financial institution. When a consumer makes an online purchase, the vendor

typically sends a confirmation email, or provides a confirmation page, that the consumer can print out.

The CE program does not currently ask respondents to provide these electronic records of expenditures. Their potential role in the CE data collection process deserves attention. These records cannot provide all of the data required by the CE program, however. They potentially offer a way to help respondents remember expenditures without a great deal of effort.

Changes in the Retail Environment

Any redesign of the CE program should take into account several recent developments in the retail landscape. These changes include the consolidation of the grocery and pharmacy retail industry; the expansion of loyalty card programs, especially among supermarket and drug store chains; the prominence of "big box stores" like Walmart and Target that offer a very wide range of merchandise; and the emergence of firms that employ data mining and data aggregation techniques in the retail industry.

In recent years, the grocery and pharmacy retail industries have greatly consolidated through mergers and acquisition. By 2005, the 20 largest U.S. grocery store retailers accounted for 61.6 percent of total sales, up from 40.6 percent in 1995 (Kaufman, 2007), while Walgreen's, CVS, and Rite Aid represented 73.2 percent of drugstore sales, and Costco and Sam's Club accounted for 89.5 percent of warehouse club sales (Fox & Sethuraman, 2006). This consolidation in the retail industry means that an increasing proportion of consumers make purchases at fewer retailers. As a result, relatively few retail chains hold data about the purchases of a large proportion of CE respondents. If those retail chains were willing to make those data available, the CE program would have a method of verifying or enhancing the data provided by the respondents.

The expansion of loyalty card programs may be relevant to the redesign of the CE program. Retailers that offer loyalty cards potentially could monitor the purchases of individual consumers. If CE respondents provided their loyalty card numbers, and retailers were willing to release purchasing data, the CE program would have access to objective information about the respondents' expenditures. Of course, this idea has some drawbacks. Consumers sometimes forget to provide their loyalty card to the cashier when they make a purchase. Some consumers may lend their loyalty cards to friends. Also, most retailers, including Walmart, have no loyalty card programs.

"Big box stores" such as Target and Walmart offer merchandise in a wide range of categories, including groceries, clothing, health supplies, electronics, and so on. CE respondents may have difficulty remembering expenditures at these outlets. For example, if a CE respondent purchased a large amount of food items and one article of clothing and one music CD during a shopping trip, the consumer may not recall the clothing item and the CD. The respondent may need to refer to a receipt to recall those items. The prominence of "big box stores" increases the need for records such as receipts to document CE respondents' purchases.

In recent years, retailers have relied increasingly on data aggregators and data mining companies such as Nielsen, SymphonyIRI, Teradata, and Catalina Marketing. These firms provide a way for the retailers to understand the purchasing behaviors of their customers. Conceivably, these firms have data that may be useful to the CE program.

Increasing Flexibility

The current design of the CE is not adequately flexible or adapted to the specific needs of the individual respondents. The procedures for collecting data in the Diary survey or the Quarterly Interview survey do not vary very much among the respondents. If the data collection methods were made more flexible, respondents might feel less burdened and more willing and able to report their data accurately and completely.

Moreover, the current CE Quarterly Interview survey is difficult to change when changes are needed. The survey is administered as a computer-assisted personal interview (CAPI). With this system, much time and effort is required to make changes to the interview procedures and questions. As purchasing methods and technologies change, the CAPI interview methods should have the flexibility to shift rapidly in response.

The CE data collection methods also could be tailored to the characteristics of the respondents. For example, some respondents may be quite comfortable with the existing paper-and-pencil Diary survey methods while others would strongly prefer to keep their data using a desktop computer or a smartphone. Some respondents may agree to download files of their credit card and checking account activity, while others may perceive a privacy threat and refuse. Some respondents may have no checking account and therefore have no checking account data files to download. In particular, lower income households disproportionately have no checking accounts and use cash or money

orders instead (Blank & Barr, 2009; Kim, 2001; McKernan & Sherraden, 2008). During economic downturns, increasing numbers of consumers stop using or terminate their credit cards (Foster, Meijer, Schuh, and Zabek, 2011).

A redesign of the CE program should take into account such individual differences. By offering respondents a range of options for reporting their expenditure data, the CE program might collect high-quality data while minimizing respondent burden.

Potential Approaches to the Redesign of the CE Program 3

This section presents some potential approaches to redesigning the CE program, and discusses the advantages and drawbacks of each approach. The redesign options are intended to address the important issues facing the CE program: measurement error, respondent burden, changes to the retail landscape, the emergence of new technologies, and the need for greater methodological flexibility.

Alternative Data Sources

Currently, the respondents are the sole source of data in the CE program. A redesigned program could take advantage of other sources of data (Tucker, 2011; Eltinge, 2011).

Data from Retailers or Data Aggregators. Some larger retailers maintain databases with sales and price information. Data aggregators obtain the databases from many of these retailers, with the goal of providing summaries that can guide marketing decisions. For example, they use the data to help manufacturers decide upon the best pricing and advertising strategies for their products. Their goal is not to study the purchasing behavior of individual households. Nonetheless, these data could be useful for at least some aspects of the CE program. For example, aggregated retail data may be useful for CPI calculations, which do not require household-level information.

Neilsen is an example of a data aggregator. According to the company's web site, Nielsen obtains sales and price data electronically from "virtually every major retail chain." Walmart, the nation's largest retailer, was late to join the Nielsen panel of retailers; it did not start to provide data to Nielsen until July, 2011 (Wohl, 2011). Nielsen also attempts to collect data about retailers that do not participate in this program. For example, before Walmart began to provide its sales and price data, Nielsen conducted interviews with Walmart shoppers to obtain those data.

Several other market research firms obtain price and sales data from retailers. For example, NPD (2011) claims to get data from a panel of 900 retailers representing 150,000 stores in and out of the United States. The data include sales per store and pricing.

The data from aggregators such as Nielsen or NPD may potentially allow the CE program to generate estimates that could be used for the CPI. The major advantage of this approach is that the data are collected without the need for any respondent interviews. However, the approach has several drawbacks.

First, these data may be seriously flawed. Perloff and Denbaly (2007) raised questions about the sales and price data that Nielsen obtains from major vendors. They said that this dataset generally does not contain data from smaller retailers, creating a systematic bias. Perhaps a more important criticism is that the methods used to collect these data are not made public.

In addition, aggregators provide little information about the statistical properties of their data. Without this information, the CE program may not adequately identify the best methods for using the data in CPI calculations.

Another major drawback to using data from aggregators is that it creates a dependency between BLS and the data aggregator. If the company changes methods, or discontinues its work, the CE data collection effort would be in jeopardy.

An alternative to using data from aggregators may be more feasible: The CE program could collect sales and price data directly from a sample of retailers and estimate expenditures for purposes of the CPI. This method is similar to the methods that the BLS uses to collect data in the Current Employment Statistics program or the Job Openings and Labor Turnover Survey (JOLTS). For the data to be useful, however, the retailers would have to provide microdata, tailored to the purposes of the CE program. The BLS might have to develop special tabulations for these microdata.

To our knowledge, the BLS has not approached individual retailers to obtain their sales and price data. Retailers may well be reluctant to provide these data. However, these retailers might seriously consider such a request from BLS, especially when they realize that the data will be used for important purposes such as calculations for the CPI. Retailers might also be motivated by a clear incentive to provide their data. For example, retailers turn their data over Nielson because of the useful information they get in return, which they use to guide their marketing decisions. If BLS could similarly develop feedback reports that summarized the information in useful ways for the retailers, they may be willing to provide their sales and price data. BLS may have to negotiate agreements with each retailer on an individual basis.

National Consumer Panel. In 2009 Nielsen and SymphonyIRI launched a joint venture, called the National Consumer Panel, with the goal of studying purchasing habits of American households (SymphonyIRI, 2011; Nielsen, 2011). Each household in the National Consumer Panel is given a small barcode scanner. Each time someone in the household makes a purchase, a respondent in that household uses the barcode scanner. First the respondent scrolls a list of retailers on the scanner screen and selects the name of the retailer, if it is on the list. Then the respondent scans the bar codes of all the UPC-labeled items that were purchased at that retailer. If the product has no UPC barcode, the respondent scans an appropriate barcode from a set of barcodes that they keep on hand. SymphonyIRI and Nielsen receive the scanned data daily from each household in the panel. According to literature from Nielsen and Symphony IRI, over 52,500 households are currently participating in the National Consumer Panel.

The goal of the program is to obtain highly accurate expenditure data using bar code scanning, rather than relying on respondents' recollections of their purchases. Nonetheless, the quality of National Consumer Panel data can be questioned.

U.S. Department of Agriculture (USDA) researchers considered and rejected the idea of using data from SymphonyIRI-Nielsen's National Consumer Panel for the National Household Food Acquisition and Purchase Survey (FoodAPS). Perloff and Denbaly (2007) explained the reasons for that decision. They wrote that the properties of the sample drawn for the National Consumer Panel have never been made public. They argued that documentation was absent on sampling, data collection procedures, self-selection, nonresponse, and attrition. Further, they wrote that "no formal statistical studies have been conducted to measure the magnitude of the size of any potential bias." They argued that the demographic makeup of the National Consumer Panel appeared to be very different from that of the overall U.S. population. They also pointed out that Nielsen and SymphonyIRI did not maintain longitudinal data, but discarded data that were more than 3 years old, making any study of trends over time all but impossible. They further pointed out that Nielsen and SymphonyIRI charged high prices for the data. Sales and price data for just one or two product categories cost hundreds of thousands of dollars. Finally, they found that Nielsen and SymphonyIRI refused to release certain parts of the dataset, such as data regarding brand market share, and data about a household's participation in programs such as WIC.

Perloff and Denbaly (2007) concluded that "relying on commercial vendors is unattractive because these firms charge very high prices, do not fully disclose the nature of their data, provide data for only very short periods, and report only variables that are important for commercial customers and not all variables that are important for researchers. One approach to ameliorating data shortages for

research would be to have government agencies or nonprofit organizations collect the ideal datasets or provide incentives to commercial providers." USDA ultimately decided to collect data itself for FoodAPS, using its own sample.

Einav, Leibtag, and Nevo (2008) also questioned the quality of the data collected using the home barcode scanning method. They compared data that respondents provided using a bar code scanner with data provided by the retailers themselves. The authors found that "…for approximately 20 percent of food-shopping trips recorded in the (bar code) data, there was no corresponding transaction in the retailer's data." Apparently for those data, the respondents entered the incorrect supermarket chain, or the wrong date of the purchases. The authors continued, "For the trips that did match up, roughly 20 percent of the (purchased) items were not recorded." Clearly, asking respondents to scan the bar codes of their purchases cannot by itself solve the challenges facing the CE program.

FoodAPS. The Economic Research Service and the Food and Nutrition Service of USDA will soon collect home barcode scanning data from a random sample of American households in FoodAPS. The respondents are asked to scan the barcodes of every food item that they buy. If the item carries no UPC bar code, the respondents scan an appropriate bar code printed in a booklet. The respondents use a very small scanner with no keyboard or screen. The respondents are also asked to keep receipts for all of their food purchases, both for food to be eaten at home and food to be eaten away from home, and to keep a diary of their food purchases on special forms. Respondents are further asked to call an interviewer after they purchase a meal away from home, so that they can report the details of the meal in a CATI interview.

The FoodAPS data include quantities, prices, and expenditures for all at-home and away-from-home foods and beverages purchased or acquired at no cost from all sources. The data also include household characteristics, including income, participation in programs such as the Special Supplemental Nutrition Program for Women, Infants and Children (WIC), and general information about non-food expenditures (U.S. Department of Agriculture, 2011).

The FoodAPS data appear to be equivalent in many ways to the data that the CE program collects for food expenditures. Conceivably, FoodAPS data could supplement CE data for food expenditures. However, the CE program probably should not discontinue collecting data about food expenditures and use FoodAPS data instead. The FoodAPS sample size of 5000 is probably too small for FoodAPS data to be comparable to CE data. Also, the FoodAPS data collection period, March to June, 2012, is not a good fit with CE methods. In addition, FoodAPS collects data about

food obtained at no cost, such as in programs that give free meals to indigent schoolchildren, while the CE program collects data only about purchased food.

Nonetheless, FoodAPS does seem to have some promise as a data collection mechanism for the CE program. With some changes to the FoodAPS data collection methods, FoodAPS data could potentially meet the needs of both USDA and BLS. Discussions between the two agencies are warranted.

Data Mining Firms. The business focus of two companies, Teradata and Catalina Marketing, is to mine retailers' databases of purchases. Teradata provides data mining software that retailers can use to study their customers' purchasing behaviors. Catalina Marketing provides data mining services to retailers. It also markets systems that print coupons for consumers when they complete their purchases. These coupons are tailored specifically to the individual consumer, based upon the consumer's purchasing history, as recorded in loyalty card data. For example, a consumer who has a history of purchasing breakfast cereals might receive a cents-off coupon for cereal redeemable only at that retailer. Catalina Marketing's web site claims that the company's systems operate in 25,000 U.S. stores, logging more than 300 million transactions per week.

The Baltimore City Paper published an article on Teradata and Catalina Marketing, covering such issues as data privacy (Jackson, 2003). The article concluded that "Companies like Catalina and Teradata use…data to spot trends and insist they have little use for individual information." The article quoted a Catalina executive as saying, "it would be difficult, though possible, to single out and review one person's buying history, but …the company has no business interest in doing such." The potential utility of the data that Teradata or Catalina Marketing could provide to the CE program is unclear. The optimum path may be for BLS to seek data directly from individual retail chains that have loyalty card programs.

Administrative Data Linked to Respondents

Many federal surveys collect data simultaneously from households and from administrative sources. The administrative data are used to verify or to supplement the data provided by the household respondents. The CE program currently collects data only from household respondents. The program does not attempt to collect any administrative data from retailers, utilities, banks, apartment

lessors, or other potential sources. However, administrative data potentially could reduce measurement error in the CE program.

The idea of obtaining data from retailers has an obvious appeal: those data are likely to be more consistently accurate than respondent-provided data are likely to be. Other federal data collection projects have used administrative data to improve the quality of data that respondents had provided from memory. The experiences of those researchers suggest that in many cases administrative sources can provide data that respondents omitted, correct data that respondents misreported, and verify data that respondents correctly provided.

For example, in the Medicare Current Beneficiary Survey, researchers asked Medicare beneficiaries about the nature and the cost of the medical services that they have received. The researchers linked the respondents' self-reported data to Medicare administrative files. By linking the data this way, 39 percent of missing cost data were corrected, and 22 percent of erroneous reports about whether Medicare was a payer were corrected (Eppig and Chulis, 1997).

In the National Immunization Survey, telephone interviewers call respondents and conduct an interview about the immunization history of children in the household. The respondents' reports are later verified; the medical providers who immunized the children are asked to provide the dates of the child's prior immunizations. The data from the medical providers have improved the quality of the respondents' reports. For example, data from the medical providers revealed that 61 percent of the children whose parents' reports suggested that their immunizations were not up-to-date actually did have up-to-date immunizations (Battaglia, Shapiro, and Zell, 1996).

The Energy Information Agency of the Department of Energy collects data about energy expenditures from a nationally representative sample of housing units in the Residential Energy Consumption Survey (RECS). The interviewers ask about household energy characteristics, such as the sources of energy used, household demographics, and energy usage patterns. When the household respondent in a rented property is unsure of the answers to the interviewer's questions, the interviewer attempts to collect the data from the rental agent. After collecting the household data, the Energy Information Agency conducts the Energy Supplier Survey (ESS), to collect administrative data. The ESS is a mail survey of energy companies that serviced the households that participated in the RECS. The ESS collects data about the amounts of natural gas, electricity, fuel oil, and propane that the RECS households used. The ESS data are used to correct and augment the RECS data.

Linking respondent reports in the CE surveys with administrative data might also bring about improvements in data quality. For example, data from retailers could contain the date of a purchase, as well as the price, the sales tax, and a description of the item purchased. The data could supplement a respondent's data, and help ensure that the respondent did not forget to report the purchases. The respondent would still have to provide additional information needed for the CE surveys, such whether the expenditure was for a person outside the household.

However, the idea of obtaining data from retailers poses considerable challenges. Retailers as a rule do not release data about the purchases of individual customers. Some retailers do have loyalty card programs and could in principle release the purchasing histories of individual customers. However, retailers do not generally release any customers' purchasing histories. The only exception is that pharmacies do usually release customers' purchasing histories for prescription drugs. Moreover, only retailers who have loyalty card programs can easily collect data for individual households. Most small retailers do not have loyalty card programs. Even Walmart, the nation's largest retailer, has no loyalty card program.

When a retailer does release loyalty card information about individual customers, it can be a newsworthy event. For example, the Denver Post (Raabe, 2010) reported that the Centers for Disease Control and Prevention (CDC) obtained information about customers' purchases from Costco to track down the source of a salmonella outbreak. CDC researchers examined the purchases of afflicted Costco customers, using loyalty card data. They identified several products that most or all of these customers had purchased. Soon the CDC researchers discovered the salami product that was to blame for the outbreak.

The Denver Post article mentioned that Safeway, King Soopers, Krogers, and Costco used loyalty card data to facilitate Class I product recalls, the category in which customers could be harmed or sickened by a faulty product. For example, Costco used loyalty card data to identify customers who purchased a toy that posed a choking hazard for children. Costco contacted these customers and warned them to discard or return the toy. Similarly, King Soopers used loyalty card data to identify customers who purchased seasonings that had spoiled. They employed an automated telephone system to warn these customers.

The needs of the CE program do not have the urgency of a dangerous product recall, or a CDC investigation of an epidemic. On the other hand, in the 1990's, Arbitron was able to persuade some retailers, including A&P, to provide point-of-sale data for the households in its consumer panel. The

retailers were motivated to provide these data because Arbitron gave them valuable market research analysis in return.

Obtaining administrative data for the CE program would not be straightforward. With other surveys, such as the Medicare Current Beneficiary Survey, or the National Immunization Survey, or the Residential Energy Consumption Survey, only a relatively few sources needed to be consulted for each household. CE households, by contrast, make purchases from a very large number of retailers and vendors. However, the BLS name has credibility. The CE and CPI programs are relatively well known. Furthermore, a comprehensive set of data, linked to respondents, is not required. Even partial data can be used for quality assessments. Many respondents may be willing to sign consents for the release of their data. The BLS might consider conducting a pilot study to investigate the feasibility of obtaining administrative data about CE households from various sources such as retailers with loyalty card programs, utility companies, and other sources. With the high degree of consolidation in the grocery, pharmacy, and warehouse club industry, data for a large number of respondents are potentially available from a relatively few sources.

Administrative Data Supplied by the Respondent

Some administrative data could be available from the CE respondents themselves. Respondents would need to expend relatively little effort to access records about their purchases made by check, electronic fund transfer, credit card, debit card, or online payment service such as PayPal. The respondents could download these data files directly from the web sites of the relevant financial institutions. Respondents who used financial software packages such as Mint could use that software to help them download those data files. Alternatively, respondents could enter their account numbers and passwords into software developed specifically for the CE program. That software could be designed using the Yodlee (2011) Software Development Kit (SDK), which automates downloads from the web sites of financial institutions. Mint and many other financial software packages were developed using the Yodlee SDK.

The data files from these financial institutions would hold a portion of the data currently required for the CE program. The files would show the total dollar amount of the respondents' purchases. In the case of the PayPal, debit card, credit card, and electronic fund transfer purchases, the files would also show the date of the purchases and the payee. The checking account files would show only the

date that the checks cleared, not the date of the purchase, and would only sometimes show the payee.

The value of the files would be that they showed, with a high degree of accuracy, the dollar amounts of the expenditures. They would also show at least the approximate dates that the expenditures occurred. The additional details that are required would have to come from a different source, such as receipts or the respondent's memory.

The data files would generally not cover, however, cash and money order transactions. Those transactions do not leave a record in a data file, except in the case of certain money orders, such as those purchased with a prepaid Western Union debit card.

However, most other expenditures would be documented in these data files. That fact raises an important question for the CE program: What proportion of consumer expenses currently are transacted using a medium other than cash or money order, and therefore leave an electronic record?

The answer to that question is suggested by a recent survey conducted by the Federal Reserve Bank of Boston. The results of the Survey of Consumer Payment Choice (Foster, Meijer, Schuh, and Zabek, 2011) reveal the proportion of consumer purchases that are transacted by cash, check, credit card and other methods. Table 1 shows that in 2009, the most recent year for which data are available, CUs reported a mean of 64.5 transactions per month. Of those, a mean of 18.4 were by cash and 0.5 were by money order. Thus, a total of 18.9 of the 64.5 monthly transactions, or 29.3 percent, were by cash or money order. These transactions would not be likely to appear in any online data file.

Table 1 reveals that a mean of 8.2 transactions per month are by check, 19.0 are by debit card, 11.2 are by credit card, 0.8 are by prepaid card, 3.3 are by online bill payment set up to occur automatically on a regularly scheduled basis, and 3.0 are by online bill payment that occurs when needed, not automatically on a regularly scheduled basis. Thus, a total of 45.5 transactions per consumer per month, or 70.5 percent of the 64.5 monthly mean, would be likely to appear in a database file that could be downloaded from the web site of a financial institution.

Table 1. Use of payment instruments in a typical month, by type of instrument (from Foster, Meijer, Schuh, and Zabek, 2011)

Number and percent of expenditures per consumer per month

	Number of expenditures			Percent of expenditures	
	Mean		Percentage change		
	2008r	2009		2008r	2009
Total payments	**67.4**	**64.5**	**-4.2**	***100.0***	**100.0**
Paper instruments	**24.1**	**26.7**	**10.8**	***35.4***	**41.8**
Cash	14.5	18.4	26.9	*20.8*	28.2
Check or money order	10.0	8.7	-13.1	*14.6*	13.5
Check	9.6	8.2	-14.0	*13.9*	12.7
Money order	0.4	0.5	23.6	*0.6*	0.8
Travelers check	0.0	0.0	-38.2	*0.0*	0.0
Payment cards	**35.4**	**30.7**	**-13.2**	***52.5***	**47.8**
Debit	21.2	19.0	-10.0	*30.8*	29.3
Credit	14.4	11.2	-21.9	*21.1*	17.3
Prepaid	*0.4*	0.8	-	*0.6*	1.2
Prepaid, per adopter†	2.6	2.6	-0.6	—	—
Electronic payments	**7.4**	**6.2**	**-15.1**	***10.9***	**9.7**
Online banking bill payment*	3.4	3.3	-3.4	*3.1*	5.1
Bank account number payment*	4.1	3.0	-26.1	*7.9*	4.6
Bank account number payment, per adopter†	7.3	5.4	-26.5	—	—
Direct deduction from income	**0.8**	**0.6**	**-31.4**	***1.1***	**0.8**

* Adjusted for changes from 2008 to 2009 in the survey definition of transaction categories.

† Estimates for this row are calculated using only adopters of a payment instrument, not all consumers. Per adopter estimate is included due to changes in the survey design that affected the rates of adoption of payment instruments, making the per consumer estimates not comparable across years.

NOTE: Superscript "r" denotes revised. Numbers in italics are not comparable across years due to changes in the survey. Numbers may not sum exactly due to rounding or missing values.

Table 2 shows that in 2009, a mean of 18.4 transactions per CU per month were for bill payments, and 2.4 of those (13.0 percent) were by cash or money order and therefore would probably not appear on any downloadable file. A total of 5.1 transactions per CU per month were for online payments. None of those were by cash and 0.1 (2.0 percent) were by money order. A total of 42.3 transactions were for retail, service, and person-to-person transactions, and 16.8 of those (40.7 percent) were by cash or money order. In summary, approximately 70.5% of all transactions could potentially lead to some type of electronic record (Figure 1).[1]

[1] The survey from which these data are derived is the *American Life Panel*. . This is a web-based panel that has recruited members from other general population surveys, including the *Monthly Survey* (MS) administered by the University of Michigan and the Face-to-Face Recruited Internet Survey Platform, administered by Stanford and ABT-SRBI (Foster, et al, 2011). Individuals who agree to be in the panel are provided access to the internet through a Web TV, if they do not have their own access. Coverage and response rates associated with a panel of this type are difficult to calculate because of the multiple stages of selection and possible attrition over time.

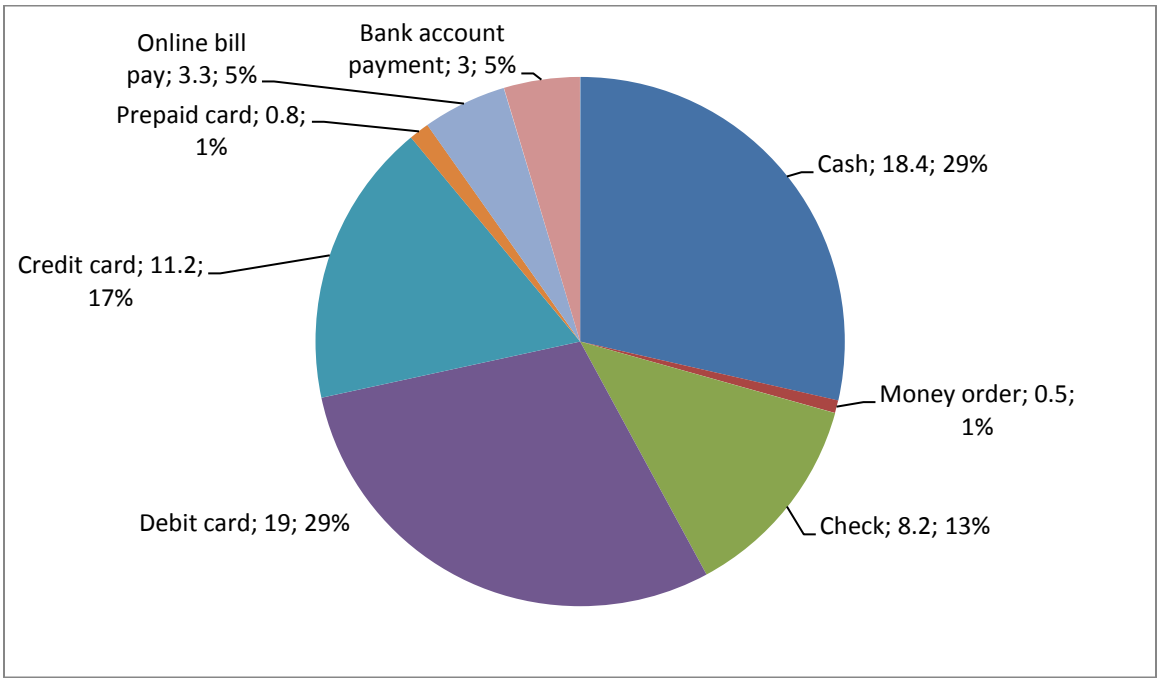
Table 2. **Use of payment instruments in a typical month, by type of transaction (from Foster, Meijer, Schuh, and Zabek, 2011)**

Number and percent of expenditures per consumer per month

Number of expenditures	Bill payments		Online payments		Retail, service and person to person	
	2008[r]	2009	2008[r]	2009	2008[r]	2009
Paper instruments	**5.1**	**5.5**	**1.6**	**1.4**	***17.9***	**20.1**
Cash	1.2	2.1	NA	NA	*13.6*	16.6
Check or money order	4.1	3.4	1.6	1.4	*4.5*	3.9
Check	3.9	3.2	NA	1.3	*4.3*	3.8
Money order	0.2	0.3	NA	0.1	*0.2*	0.2
Travelers check	NA	NA	NA	NA	NA	NA
Payment cards	**7.9**	**7.1**	**3.7**	**3.0**	***23.9***	**20.8**
Debit	4.6	4.5	2.1	1.8	*14.8*	13.1
Credit	3.6	2.6	1.6	1.2	*9.4*	7.6
Prepaid	0.0	0.1	0.1	0.1	*0.3*	0.5
Electronic payments	**6.0**	**5.2**	**1.4**	**0.7**	**NA**	**0.4**
Online banking bill payment	2.2	3.1	NA	NA	NA	0.2
Bank account number payment	4.0	2.1	1.4	0.7	NA	0.2
Direct deduction from income	**0.8**	**0.6**	**NA**	**NA**	**NA**	**NA**
Paper instruments	**7.4**	**8.3**	**2.3**	**2.2**	**25.7**	**31.3**
Cash	1.5	3.0	NA	NA	19.3	25.2
Check or money order	5.9	5.3	2.3	2.2	6.4	6.1
Check	5.6	4.9	NA	2.0	NA	5.8
Money order	0.3	0.4	NA	0.2	0.3	0.3
Travelers check	NA	NA	NA	NA	NA	NA
Payment cards	**11.6**	**11.0**	**5.5**	**4.6**	**35.4**	**32.2**
Debit	6.4	6.8	3.0	2.7	21.3	19.8
Credit	5.1	3.9	2.3	1.8	13.6	11.6
Prepaid	0.0	0.2	0.1	0.2	0.5	0.8
Electronic payments	**9.0**	**8.0**	**1.9**	**1.0**	**NA**	**0.6**
Online banking bill payment	3.1	4.8	NA	NA	NA	0.3
Bank account number	5.9	3.3	1.9	1.0	NA	0.3
Direct deduction from income	**1.1**	**0.8**	**NA**	**NA**	**NA**	**NA**

NOTE: Superscript "r" denotes revised. Numbers in italics are not comparable across years due to changes in the survey. Numbers may not sum exactly due to rounding or missing values. The notation "NA" indicates that the estimate is not available.

Figure 1. Transactions per consumer per month



In summary, these findings suggest that fully 70.5 percent of consumer purchases are documented in electronic format. More specifically, 87.0 percent of bill payments, 98.0 percent of online payments, and 59.3 percent of retail, service, and person-to-person transactions are documented in electronic purchase confirmations or in files that can be downloaded from the web sites of financial institutions.

The trend to the use of electronic methods of payment is increasing at a very fast pace. In 1995 the percentage of individuals who participated in electronic banking was approximately 4 percent, compared to 53 percent in 2007. The percentage using pre-authorized electronic fund transfers has doubled during this time period from 25 percent to 49 percent and the percentage using debit cards has more than tripled from 20 percent to 71 percent (Table 3, which is Table 1184 from the 2011 Statistical Abstracts [Census Bureau, 2011]). New technological innovations are continually emerging, such as smartphone apps for purchasing and e-receipts (Clifford, 2011), in which retailers such as Old Navy email receipts to the consumer rather than print out paper receipts. Certainly, even more forms of electronic purchase documentation will appear in the future. Any redesign of the CE program must take these innovations into account.

Table 3. Percentage of households using selected electronic banking technologies: 1995 to 2007 (from Census Bureau, 2011)

Technology	Survey of Consumer Finances					Reuters-University of Michigan Surveys of Consumers		
	1995	1996	2001	2004	2007	1999	2003	2005
Electronic								
Direct deposit of any type	53	67	71	75	80	65	70	77
ATM Card	35	55	57	65	76	59	65	69
Debit Card*	20	37	50	62	71	NA	54	62
Preauthorized debits	25	40	43	50	49	31	46	57
Automated phone system	NA	26	22	20	25	40	44	46
Computer banking	4	7	19	34	53	10	32	51
Smart card**	1	2	3	NA	NA	NA	6	12
Prepaid card†	NA	NA	NA	NA	NA	NA	73	73
Nonelectronic								
In person	87	81	78	78	85	NA	NA	NA
Mail	59	55	51	51	59	NA	NA	NA
Phone (talk in person)	NA	43	42	42	57	NA	NA	NA

Source: Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin*, July 2009, and unpublished data.

* A debit card is a card that automatically deducts the amount of a purchase from the money in an account.

** A smart card is a type of payment card containing a computer chip which is set to hold a sum of money. As the card is used, purchases are subtracted from that sum.

† Prepaid cards are cards that contain a stored value, or a value that has been paid up-front, allowing you to use the card much like cash. As you use the card, the prepaid value is drawn down. Examples are phone cards and gift cards. Smart cards are different from prepaid cards in that you can add money to the card at special machines designed for smart cards or sometimes at ATMs.

NOTE: Covers only those households that access services (other than by check or credit card) at a bank, thrift institution, or credit union based on sample surveys. For details on the Survey of Consumer Finances, see the Federal Reserve Board, <www.federalreserve.gov/boarddocs/surveys\>. The Reuters/University of Michigan Surveys of Consumers is based on data from approximately 1,000 respondents. For details, see the University of Michigan Survey Research Center, <http://www.sca.isr.umich.edu/>.

Records of Expenditures

One of the most obvious ways to improve the quality of CE data and to reduce the respondents' reliance on memory is to encourage the respondents to keep documentation of their expenditures. This documentation could be in the form of receipts, including paper cash register receipts and printouts of online purchase confirmations. If no receipt was available, such as for purchases from vending machines, the documentation could be in the form of a notation; the respondents could create a record of the purchase using a smartphone or simply by writing down information about the purchase on a short paper form.

Evidence from several federal surveys suggests that when respondents keep records of some kind, they provide better-quality data than they do when they simply try to recall the data. Such records could be very simple, such as receipts or notations.

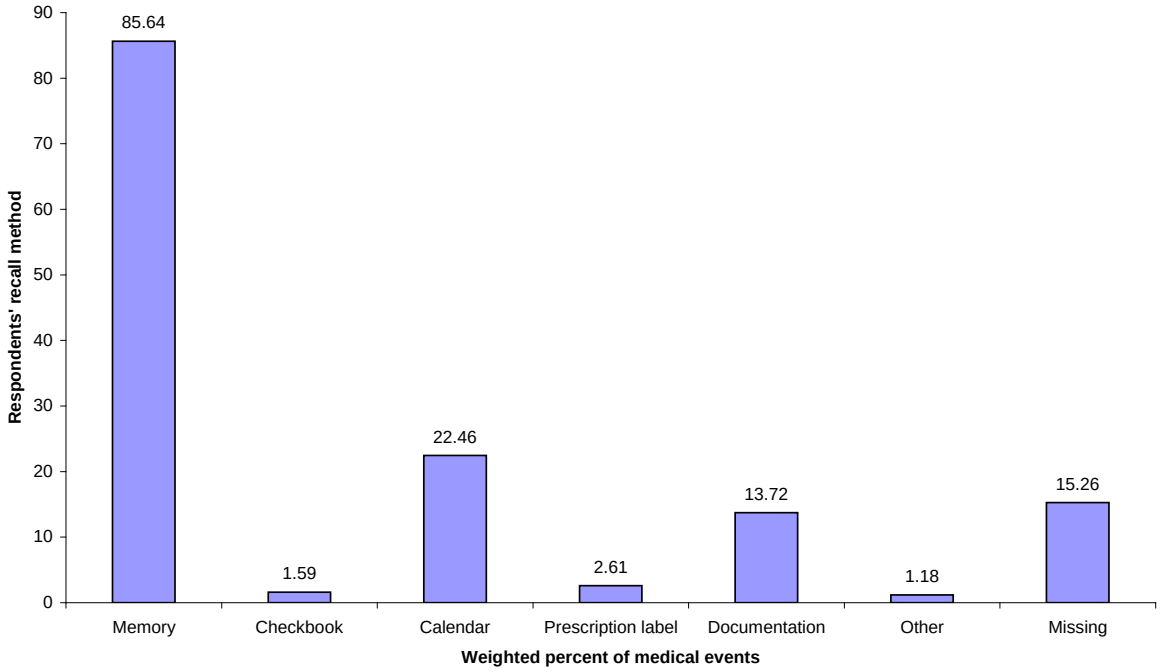
Medical Expenditure Panel Survey. For example, data from the Medical Expenditure Panel Survey (MEPS) suggest that respondents who rely on memory alone to report their data are more likely to err than are respondents who use a memory aid. All the respondents in MEPS receive a calendar to record their medical expenditures. They are asked to record each health-related product or service they obtained, how much it cost, and the vendor who provided it. The calendar has a pocket for the respondents to keep receipts, explanation of benefits forms, or bills to document their medical expenditures.

MEPS respondents, however, often failed to heed the instructions to keep up the calendar and to save receipts or other documentation. For 59.8 percent of the medical expenditures reported in 2003, the respondents relied solely on memory to report out-of-pocket medical expenses. For only 20.0 percent of the medical expenditures did the respondents use the calendar and for only 9.7 percent did the respondents refer to documentation such as receipts to help them report out-of-pocket medical expenditures (Kashihara and Wobus, 2007).

When respondents did use the calendar or keep documentation, their data tended to be more accurate than when they relied on memory alone. The respondents' reports were compared with medical providers' records, which were assumed to be accurate. Use of the calendar improved the accuracy of respondents' reports for out-of-pocket medical expenditures. In a logistic regression analysis, the odds ratio of 1.16 was significant ($p<.05$), suggesting that use of the calendar was associated with accurate reports, as compared with the use of memory alone. Use of documentation such as bills and explanation of benefit forms improved the accuracy of respondents' reports on insurance payments. The odds ratio of 1.30 was significant ($p<.05$), suggesting that the use of bills was associated with accurate reports (Kashihara and Wobus, 2007).

More recent MEPS data, for 2004 to 2006, are shown in Figure 2. For only 22.5 percent of the medical expenditures did the respondents keep notes on a calendar. For only 13.7 percent of the medical expenditures did the respondents refer to documentation such as receipts. Variables which may be proxies for low socioeconomic status, including lower family income, less educational attainment, Medicaid coverage, and lack of health insurance, were associated with a decreased tendency to record medical data on the calendars, or to keep documentation, and an increased tendency to rely on memory to report data. In addition, demographic variables such as African-

Figure 2. Weighted percent of medical expenditures for which the respondents used each recall method (from Schneider, Gustavson, Li, Göksel, DiGaetano, and Edwards, 2009)



NOTE: Interviewers could note that respondents could use more than one recall method for any medical expenditure. The recall method was classified as "missing" when interviewers did not note any recall method.

American race, Hispanic ethnicity, younger age, male sex, and worse health were associated with decreased use of a calendar or documentation, and increased use of memory (Schneider, Gustavson, Li, Göksel, DiGaetano, and Edwards, 2009).

The MEPS data suggest that records can improve the quality of the data obtained through respondents' reports, but respondents tend not to be as diligent as researchers might like in keeping records.

National Immunization Survey. In the National Immunization Survey, when interviewers telephone the sampled households, they ask the respondents if they have a "shot card" showing the child's past immunizations. When the respondents have a shot card, the interviewer asks them to refer to it to report the data.

Respondents who referred to shot cards tended to provide more accurate immunization data than did respondents who relied solely on memory. A total of 90.5 percent of children who appeared to be up-to-date from a shot card were actually up-to-date. Only 78.0 percent of children who appeared

to be up-to-date based on reports from the parent's memory were actually up-to-date (Brick, Kalton, Nixon, Givens, and Ezzati-Rice, 1999; Ezzati-Rice, Zell, Massey, and Nixon, 1996).

However, the shot cards introduced a systematic bias. Very rarely did a shot card show an immunization that actually did not occur. When an error appeared on a shot card, it was almost always the omission of an immunization. Thus, while errors in reports from memory might include both inaccurate reports of immunizations and inaccurate omissions, errors on shot cards included almost exclusively inaccurate omissions.

The CE program might similarly benefit if respondents could be motivated to keep records of their expenditures. The experience of the Census Bureau in the Survey of Income and Program Participation (SIPP) suggests that several techniques may be effective in motivating respondents.

SIPP. Census Bureau researchers attempted to motivate some of the SIPP respondents to use their personal records, rather than just their memory, to report their income (Bogen, Moore, and Marquis, 1994; Marquis, 1995; Moore, Marquis, and Bogen, 1996).

The SIPP researchers employed these techniques to encourage the respondents to use their records:

- Interviewers told the respondents that personal record use was the norm. They said that it was the usual, expected part of the respondent's task.
- Interviewers asked respondents who could not produce a record of income to replace it. For example, respondents who said that they had discarded their pay stubs were asked to get replacement stubs from their employers. The interviewers then called or returned to the respondents to collect the income information from the replacement pay stubs.
- Interviewers trained the respondents in recordkeeping. They showed the respondents how to store documents such as pay stubs, and how to make a written record of income that came with no documentation, such as public assistance payments.
- Interviewers telephoned respondents between interviews to remind them to keep records of all income.
- Interviewers permitted respondents to retrieve their records during an interview, even when this was time consuming.

With these procedures, the researchers achieved a substantial increase in the use of records. They compared respondents who were encouraged to use records to report their income with control respondents, who did not receive such encouragement. They found that only 25 percent of the

control respondents referred to records in the SIPP interviews. However, 87 percent of the respondents who received the intervention referred to records. The researchers described that level of record usage as "astonishingly high."

The use of records was generally associated with improved data quality. The SIPP respondents were more accurate ($p<.05$) in reporting their level of income when they referred to records. However, the intervention did not bring about any significant change in the number of income sources reported.

The intervention was associated with two drawbacks. First, it brought about a decrease in the response rate. In one wave of data collection, the control group response rate was 94 percent while the intervention group response rate was 82 percent. In the second wave, the difference persisted. The control group response rate was 98 percent while the intervention group response rate was 90 percent. This discrepancy may be attributable to the increased respondent burden associated with retrieving and reading documentation.

In addition, the per-case cost of data collection was markedly different between the control and intervention groups. While the control group costs were in the $18 to $24 range, the intervention group costs were much higher, in the $49 to $50 range.

In summary, the results of other federal surveys strongly suggest that data quality improves when respondents keep records. However, encouraging respondents to keep records poses a considerable challenge. The experience of SIPP shows that respondents can be effectively motivated. However, the effort to motivate the respondents drives up costs and causes the less cooperative respondents to withdraw from the survey.

The CE program might adopt new procedures to encourage respondents to keep receipts and other records diligently. Those procedures would have to be carefully crafted so that they did not add an unacceptable burden to the interviewers and thereby drive up costs. The procedures would also need to take into account those respondents who are not amenable to keeping records under any circumstances.

Records Created by the Respondents

CE respondents sometimes cannot refer to any record of an expenditure because they neglected to obtain a receipt, or they misplaced the receipt, or no receipt was provided, such as with a purchase from a vending machine. For these kinds of purchases, the respondent could create documentation. The purpose of the documentation would be to remind the respondent about the expenditure. That is, the documentation could help ensure that the respondent remembered to report the expenditure, and also provide at least a few details about the expenditure.

That documentation could be in any of the following formats:

- A small paper form on which the respondent writes a description of the purchase and the amount;
- A text message containing a description of the purchase and the cost;
- An email message containing a description of the purchase and the cost;
- Responses to questions on a smartphone screen about the nature and the amount of the purchase;
- Voice messages left on an Interactive Voice Response (IVR) system; the respondent calls the IVR system, which asks the respondent to describe the purchase and state the amount paid (similar to theexpensetracker.com);
- Voice memorandums recorded via a smartphone app; the respondent records a description of the purchase and the amount paid;
- Voice memorandums recorded on a digital recorder; the respondent records a description of the purchase and the amount paid;
- A photograph that the respondent takes with a smartphone; the photograph could be of the purchased items, or the store where the purchase occurred; or
- A photograph that the respondent takes of the bar code on a purchased item using a smartphone.

Many of these ideas involve a smartphone. Respondents potentially could use smartphones to fill out a short form about a purchase, or record a voice message describing a purchase, or even photograph a purchase or a bar code on a purchase.

Smartphones

All CE respondents potentially could be asked to obtain and keep receipts. Many respondents might be comfortable using an IVR system or a voice recorder or even a portable bar code scanner. However, only respondents who have and are comfortable using smartphones would wish to collect data with them.

The Pew Internet and American Life Project (Smith, 2011) conducted a telephone survey about mobile device ownership in April and May, 2011 with a sample of 2,277 U.S. adults. Respondents were contacted by both landline and cell telephone. The results indicated that 83 percent of American adults own a cell phone of some kind, and for 42 percent of them, that phone is a smartphone. Thus, the results suggest that 35 percent of American adults own a smartphone (Table 4).

Table 4. Smartphone ownership and Internet use summary (from Smith, 2011)

% of smartphone owners, cell owners and all adults who…

	% of smartphone owners who…	% of all cell owners who…	% of all adults who…
Own a smartphone	100%	42%	35%
Use the internet or email on Smartphone	87	36	30
Use smartphone to go online on a typical day	68	28	23
Go online mostly using Smartphone	25	10	8

Source: The Pew Research Center's Internet & American Life Project, April 26 – May 22, 2011 Spring Tracking Survey. n=2,277 adult internet users ages 18 and older, including 755 cell phone interviews. Interviews were conducted in English and Spanish.

Moreover, the Pew data suggest that most smartphone owners tended to make use of the device's Internet capabilities. Table 4 shows that 87 percent of smartphone owners use their device to access web sites or for email.

The Pew data suggest that certain demographic characteristics are associated with smartphone usage. Table 5 shows the relationship between smartphone ownership and demographic background. Respondents with a household income over $75,000 were more likely than others to report that they owned a smartphone. Smartphone ownership was also associated with having a college degree, age less than 50 years, African American race, and Hispanic ethnicity.

Table 5. The demographics of smartphone ownership (from Smith, 2011)

% of U.S. adults within each group who own a smartphone

All adults	**35%**
Gender	
Men (n=973)	39
Women (n=1304)	31
Age	
18-29 (n=337)	52
30-49 (n=581)	45
50-64 (n=659)	24
65+ (n=637)	11
Race/Ethnicity	
White, non-Hispanic (n=1637)	30
Black, non-Hispanic (n=261)	44
Hispanic (n=223)	44
Household Income	
Less than $30,000 (n=671)	22
$30,000-$49,999 (n=374)	40
$50,000-$74,999 (n=276)	38
$75,000+ (n=444)	59
Education level	
No high school diploma (n=229)	18
High school grad (n=757)	27
Some college (n=525)	38
College+ (n=746)	48
Geographic location	
Urban (n=618)	38
Suburban (n=1113)	38
Rural (n=465)	21

Source: The Pew Research Center's Internet & American Life Project, April 26 – May 22, 2011 Spring Tracking Survey. n=2,277 adult internet users ages 18 and older, including 755 cell phone interviews. Interviews were conducted in English and Spanish. "Smartphone ownership" includes those who say their phone is a smartphone, or who describe their phone as running on the Android, Blackberry, iPhone, Palm or Windows platforms.

These results suggest that about one-third of CE respondents might consider using a smartphone to collect their expenditure data. CE respondents with certain demographic characteristics would be more likely than other respondents to wish to use a smartphone.

However, the CE program probably cannot feasibly offer smartphones to respondent as incentives. Smartphones are costly. Their cost is greatly reduced when purchased along with a 2-year contract, but respondents may not wish to commit to paying for service over two years. Also, respondents who are unfamiliar with smartphones may require training to use them for the CE surveys. Some may discover that they do not like to use them. For that reason, only CE respondents who already own smartphones could feasibly use the devices to keep a record of their expenditures.

Number of Persons Interviewed in the Household

The current CE surveys collect data from one household respondent. However, that procedure has a disadvantage: The respondent has the burden of reporting the expenditures of all members of the CU. The respondent may be unaware of, or reluctant to report, the expenditures of other people in the household. Asking all members of the household to provide data may overcome this problem. There are several surveys in other countries that distribute diaries to multiple household members (e.g., Grootaert, 1986; Filippucci & Ferrante, 1997). In the United States, the FoodAPS survey will ask all household members to provide data.

BLS and the Census Bureau have evaluated a method in which multiple CE household members kept diaries (Westat, 2005; Edgar et al., 2006). With that method, a principal respondent was responsible for key household expenditures like groceries and utilities. All members of the household, including this principal respondent, reported on their own individual purchases. The results of the evaluations suggested that a significant number of these individual household members did report expenditures.

For example, the larger study conducted by the Census (Edgar et al., 2006) found that a significant number of the individual diaries were filled out by young males, a population that is typically not well represented in surveys. In 362 of the 591 households in the multiple respondent group, all the adults reported expenditure data. A comparison group was formed using a sample from the ongoing CE Diary survey. Overall, the distribution of expenditures across four broad categories (food at home, food away from home, clothing, other) did not significantly differ between the multiple and single respondent procedures. The number and level of expenditures was quite a bit higher for the

multiple diary survey. For example, the mean number of reported expenditures per CU was 53.8, compared to 37.8 for the production survey. Perhaps more suggestive is a comparison of the mean level of expenditures. The single diary CUs had a mean of $856 for the first week of collection compared to $1137 for the multiple diary procedure. This is a ratio of about .75. A slightly lower ratio of .65 was found for the week 2 ($838 versus $1272). The means for the multiple-diary group may have been inflated because blank diaries were not consistently included in the computations. Therefore, the differences between the mean reported expenditures of the single and multiple diary groups may have actually been somewhat smaller.

Qualitative evidence—feedback from the field interviewers—indicates that the multiple-respondent procedure seemed to encourage the principal respondent to communicate more with others in the household. The multiple-respondent procedure seemed to impress upon the principal respondent the importance of the task of collecting as many of the CU expenditures as possible.

The multiple-respondent procedure did lead to a lower response rate (61% versus 77%). However, it is unclear whether this led to significant bias. Perhaps more troubling is that it took almost twice the effort to complete a CU for the multiple diary procedure (5.1 visits versus 2.7 visits). It is difficult to know exactly how much of an increase in actual costs this implies without more information on the amount of additional time field interviewers spent completing interviews. Nonetheless, taken at face value, and based on crude interpolation from Westat cost data, this additional interviewer time could lead to up to twice the field costs of the current procedures.

The multiple-respondent procedure is worth considering because of its potential to directly address measurement error. While the Census study was not designed to draw quantitative conclusions, it does suggest that underreporting could be reduced if multiple respondents report data in each household. The multiple-respondent procedure has the potential to reduce the overall mean square error. In addition, it also may reduce respondent burden by relieving the principal respondent of having to proxy for others in the household.

Respondent Incentives

A vast literature exists on the effectiveness of incentives in boosting the response rates in various kinds of surveys. Researchers have studied incentives such as cash, checks, gift cards, and contests.

The results suggest that under appropriate conditions, carefully chosen incentives can be a cost-effective means to increase response rates.

Both Berlin et al. (1992) and RTI (2002) tested different levels of incentives for an in-person household interview. Both found that incentives increase response rates and save enough money to pay for themselves. Similarly, significant incentive effects have been found in comparable experiments involving SIPP and the Survey of Program Dynamics (Creighton, King, & Martin, 2007; Martin, Abreu, and Winters, 2001), the American Time Use Survey (Piskurich et al., 2001), and many surveys prior to those (Church, 1993; Singer, 2002).

More recently, Hicks, Connor, Ward, et al. (2009) offered respondents cash incentives in the Medical Expenditure Panel Survey. They found that incentives of $50 and $70 improved the response rates relative to the existing $30 incentive. This effect was apparent in most subgroups of respondents, including low-income, Hispanic, and Black respondents. Moreover, the incentives did not greatly increase the costs per case.

Perhaps more relevant for the CE, Goldenberg, McGrath, and Tan (2009) gave randomly selected households in the Quarterly Interview survey either a $20 or a $40 prefilled debit card prior to the first of the five interviews. The results suggest that the incentives boosted the response rates and this effect persisted at least through the fourth interview. The results further suggested that "…incentive recipients performed better when compared to the control group on most of the indirect data quality measures: they answered more expenditure questions, they used records more frequently, they provided fewer don't know and refused answers, and their reported data required fewer imputations and allocations. The $40 incentive performed better than the $20 incentive on 7 of 11 measures, but some of the differences were very small and most were not statistically significant." (Goldenberg, McGrath, and Tan, 2009).

Recommended Design: Diary Survey 4

This section presents our recommendations for redesigning the CE surveys. Our recommended redesign maintains separate Diary and Quarterly Interview surveys. The Quarterly Interview survey would collect information primarily about larger expenditures, while the Diary survey would collect data primarily on smaller, more frequent expenditures.

Overview

In our recommended redesign, the Diary survey would continue to require the respondents to record all their household expenditures over a 2-week period. The methods that each respondent use to keep an expenditure diary would be tailored to the needs and preferences of the respondent. Some may use the current method involving a paper diary. Some may save paper receipts and mail them to a central data repository. Some may save paper receipts, scan them, and send the scan files to the data repository electronically. Some may also download financial data files, such as for credit card accounts, and send those files to the data repository electronically. Some may instead mail in photocopies of the monthly statements for their credit card, checking, and other accounts.

Our recommendations for the Quarterly Interview survey call for four contacts with the household over a 15-month period. Three contacts would be face-to-face, and one would be by telephone. Additional telephone contacts would be made to motivate the respondent to keep track of expenses. In the first face-to-face interview, the interviewer would collect information about the household, obtain consents to release records, select the methods by which the respondent would gather and report data, and conduct a bounding interview.

Over a 3-month period, the respondent would report household expenditures. The methods that each respondent would use would be tailored to the needs and preferences of the individual respondent. The respondents in the Quarterly Interview survey could save and mail in their paper receipts, or they could save, scan, and send the receipts to the repository electronically. They could download the financial data files and send them to the repository electronically, or they could mail in photocopies of their checking, credit card, or other monthly statements.

During this 3-month period, the interviewer would call the respondent three times: 2 weeks, 1 month, and 2 months after the initial interview. The interviewer would ask if the respondent had any questions or was experiencing any problems. The interviewer would ensure that the respondent understood the data collection procedures, and try to resolve any issues that might be keeping the respondent from following those procedures. The purpose of these calls would be to keep the respondents motivated to report their data.

The second face-to-face interview would occur 3 months after the first. The nature of this interview would depend on how completely the respondent provided expenditure data over the 3-month period. The interviewer would collect information to augment the data that the respondent had reported.

One year after the initial interview, the respondent would again keep and report expenditure data for a 3-month period.

In both the Diary survey and the Quarterly Interview survey, CE field representatives would attempt to obtain purchasing history data from certain retailers. Agreements may be in place for retailers to provide these data using the loyalty card numbers of the respondents.

A summary of the differences between the current procedures and the recommended procedures is shown in Table 6. The current Diary survey and our proposed redesigned Diary survey both employ a 2-week data collection period, with two 1-week diaries. However, our recommended redesign emphasizes the use of both electronic and paper records, uses multiple respondents in each household, and collects data from administrative records.

Our recommended redesigned Quarterly Interview survey differs in many ways from the current design. The current design is a panel survey with four data collection waves spaced 3 months apart, using a combination of in-person and telephone interviews. The redesigned survey reduces this to two waves, spaced 12 months apart. There are three in-person interviews and one telephone interview. This difference reduces burden by requiring only two waves of data collection, with more time in between. As with the Diary survey, there is a greater emphasis on the use of electronic and paper records. There is also an incentive and an administrative survey to substitute for collecting data directly from the respondent.

Table 6. Comparison of design features of current and redesigned surveys

Design feature	Current collection	Redesigned survey
Diary Survey		
# of Recall Interviews/CU	0 – 2 (depends on use of diary)	0–2 (depends on record use)
Reference period	1 week	1 week
Sample Design	Cross sectional	Cross sectional
Mode	Paper	Web and paper
In-person/telephone interviews	3	1–3, depending on use of records
Telephone reminders	Yes, unknown how many	1–4 depending on use of records
Number of respondents	1 person	All persons 14+
Incentives	None	Monetary and nonmonetary
Personal electronic records	None	All respondents willing to use it
Use of receipts	Yes, not strong emphasis	Yes, very strong emphasis
Administrative Record Survey	No	From retailers used by respondents
Interview Survey		
# of Recall Interviews/CU	4	2
Reference period	3 months	1, 3, 12 months depending on item
Sample Design	Panel – 4 waves (3 months apart)	Panel, 2 waves (12 months apart)
Mode	CAPI	Web, CAPI
In-person/telephone interviews	5	3
Telephone reminders	Not known	6–7, depending on use of records
Number of respondents	Main CU respondent	Main CU respondent
Incentives	None	Monetary and nonmonetary
Personal electronic records	None	All respondents willing to use it
Use of receipts	Yes, not strong emphasis	Yes, very strong emphasis
Administrative Record Survey	No	From retailers and Utility Companies

This chapter and the succeeding chapter respectively provide the details of the redesigned Diary and Quarterly Interview surveys.

Diary Survey Procedures

Our recommended redesign methods for the Diary survey are tailored to the preferences and the technological abilities of the respondents. Some respondents have no interest in using any of technology and would prefer to keep paper records, while others will wish to use technology.

Respondents in the Diary survey would report their expenditure data over a 2-week period. The schedule and procedures are shown in Table 7. The procedures would consist of 5 total contacts by the interviewer: (1) initial contact, (2) 3-day reminder call, (3) 1 week contact to finalize purchases for prior week, (4) 10 day reminder call and (5) final contact to finalize purchases over the previous week. Table 7 provides an outline of the mode for each of these contacts and an outline of the

procedures for each one. Below we describe in more detail how the data would be reported by the respondent and the interviewer's role.

Table 7. Schedule for the redesigned Diary survey

Initial interviewer visit to the household – In person ■ Respondent signs consent form to obtain information electronically ■ The method that respondent will use to provide data is chosen (paper or electronic) ■ Interviewer provides a scanner to respondents who will report electronically ■ Interviewer provides an envelope or box for respondents to store receipts ■ Interviewer provides diary forms to respondents who will neither mail nor scan receipts
2-3 days after initial visit – Telephone ■ Interviewer calls respondent, helps respondent overcome any barriers to providing data ■ Interviewer continues to monitor respondent's data reporting activity
7 days after initial visit – Telephone/In person ■ If the respondent appears to be sending data electronically or on paper diligently, interviewer telephones; otherwise interviewer visits in-person ■ If interviewers visit in-person administer a recall interview ■ Interviewer continues to monitor respondent's data reporting activity
10 days after the initial visit – Telephone ■ Interviewer calls respondent, helps respondent overcome any barriers to providing data ■ Interviewer continues to monitor respondent's data reporting activity
14 days after the initial visit – Telephone/In person ■ If the respondent appears to be sending data electronically or on paper diligently, interviewer telephones; otherwise interviewer visits in-person ■ Interviewers who call close out the data collection ■ Interviewers who visit in-person administer a recall interview and close out the data collection

Respondents Who Elect to Submit Their Expenditure Data Electronically. The system that Westat proposed in the BLS project *Data capture technologies and financial software for collecting consumer expenditure data* is the model for the proposed Diary survey. The redesign is based on the concept of a central repository. The central repository receives, stores, and manages expenditure data from all the respondents. Each respondent has his or her own email address in the repository, so that the respondent can provide data electronically.

File Downloads. Respondents periodically download the data files for their credit card, bank, and PayPal accounts and email them to the repository. They could do this in either of two ways. First, software could be developed for this purpose. This software would download files much like existing financial software products like Mint does: Respondents select their financial institution from a list, enter their user names and passwords, and the file download occurs automatically. The software then would automatically email (or transfer via FTP) the files to the repository.

Some respondents might wish to have greater control over the process of downloading these files and emailing them to the repository. They may fear that the privacy of the files could be compromised. For that reason, respondents could choose to download the files directly from the web sites of the financial firms and email them to the central repository themselves.

The data in credit card and PayPal files show the date, amount, and vendor for each purchase. The data in bank account files show the amount for purchases by check and the date that the check cleared. The data in bank account files show the vendor, amount and date for purchases by debit or automatic fund transfer.

Receipts. Respondents would retain all receipts for their purchases. They would be given stickers to put on their wallets reminding them to ask for receipts. When a receipt was in the form of an online purchase confirmation, they would print it out.

The respondents would scan their receipts and email the scan files to the repository. Respondents could be given small receipt scanners (similar to the NeatDesk digital scanner) and special software for this purpose. The software would automate much of the process. It would instruct the respondent how to use the scanner, prompt the respondent to start scanning, and automatically email (or transfer via FTP) the scans to the repository.

The repository would receive the scans of the receipts and convert them to text using Optical Character Reader (OCR) software. From this text, the system would derive information about each item purchased. Software could be developed to extract the individual items that comprise the total purchase. For example, if a respondent purchased ten items at a supermarket, the software would capture the description and the cost of each of the ten items listed on the single receipt. Creating software with this capability will be challenging. Receipts often contain a great deal of text other than a description of the purchase and the cost (e.g., "1.5 lb @ $2.00/lb", "Bonus card savings 10%"). The software would need to ignore that extraneous text and locate the text that describes the individual items purchased and the corresponding costs (e.g., Apples $4.22). The software could also be designed to identify returns and exchanges listed on a receipt.

Several software products already exist that extract data from a document. The AnyDoc suite of software products and Nuance OmniPage can extract data from documents. However, they lack the functionality to extract expenditure data from receipts. New software would need to be created to extract data about purchases, returns, and exchanges from the text of scanned receipts.

Having identified the individual items purchased, the system would then attempt to classify each purchase according to the CE categorization scheme. For example, the system would recognize that "cheese" was a food item and "shirt" was a clothing item. The system could do this by consulting a database of expenditure terms created specifically for the CE program.

Data Matching. The system could then match the receipt data with data from the credit card, checking, PayPal, and other financial files. For example, a receipt might show a purchase of $5.52 at Home Depot on July 4, 2011. A credit card data file might show a purchase of the same amount, at the same store, on the same date. The system would categorize these data as a match. In this way, the system would recognize that the data from the two sources both correspond to a single purchase.

This expenditure matching procedure is not novel. It is a standard feature in financial software packages such as Mint (2011) and Moneydance (2011).

The data on receipts for purchases by check, credit card and PayPal will match the corresponding data in the credit card, checking, and PayPal files. Data on receipts for other kinds of purchases, such as cash purchases, will not match any data in those files.

In addition, some of the purchases documented in the bank, credit card, and PayPal files may not match any receipt data. For example, if a respondent did not obtain and scan a receipt for a purchase, the system will not be able to match the data in the check, credit card, or PayPal file for that purchase with any receipt data. Also, an electronic fund transfer from a bank account, such as for a loan payment or a utility bill payment, will not have a corresponding receipt. Therefore, those expenditures will appear in the data file for a financial account but the system could not match the expenditure with any receipt data.

Expenditures with No Receipt and No Electronic Record. The respondents will sometimes make purchases that have no receipt and that have no electronic record. For example, expenditures transacted with cash or money order typically have no electronic record. Some expenditures such as vending machine purchases involve no receipt. Respondents would need to keep a paper-and-pencil record of these expenditures. They could be given a rule: If you are not given a receipt, create your own.

This record could be in the form of a short form that the respondents complete. One possibility is that the respondents complete one form per purchase. They could then scan the form along with the receipts and email the scan to the central repository.

Web Questionnaire. The data in the financial data files, receipts, and short forms would not be sufficient for the CE program. Table 8 shows the kinds of data available from the various sources. The table shows that receipts have most but not all of the information required by the CE program.

In order to collect the necessary information, we recommend that the repository system periodically ask each respondent to complete a web questionnaire. The repository system would create the web questionnaire automatically and email it to the respondent. Each page of the questionnaire would ask about a single item purchased.

For example, the web questionnaire might pose a question like "On July 4, 2011, you purchased three items at Home Depot. One item was a hammer for $8.95. Was this item for someone on your household list?" The system would consult the database of terms and categorize "hammer" as "other" rather than "food and drinks" or "clothing."

The following examples illustrate how the repository could handle various purchases.

Credit Card Purchase with Receipt. Respondent A buys 20 items at Safeway, including 3 pounds of apples for $6.00, on July 6, 2011, paying for the purchase by credit card. Respondent A scans the receipt for this purchase. The scan goes to the repository. The system applies OCR to capture the 20 items on the receipt, including the apples. The system matches the receipt with the corresponding charge in the respondent's credit card file. Later the system sends an email that asks Respondent A to complete a web questionnaire. One of the pages in that questionnaire contains this text: "On July 6, 2011, you purchased apples for $6.00." The page also shows the relevant lines from the scanned Safeway receipt, which might read "Apples……$6.00." The page then presents questions to elicit the information required by the CE surveys. The system could consult the database of expenditure terms to recognize that "apples" is categorized as food. The web survey would ask CE questions appropriate for a food, for example, whether it is for consumption at home, and when the respondent replies "yes," whether it is for someone outside the household.

Table 8. Information that appears on various kinds of records of expenditures

	Date of trans-action	Service dates	Total price	Tax amount	Tax rate	Shipping fees	Item verbatim descrip-tion	Item descrip-tion	Item ID number	Item price	Item rebate price
Cash, no receipt											
Paper receipt	U		U	U	S	U	U		S	U	U
Electronic confirmation	U	U	U	U	S	U	U		U	U	U
Downloaded file: EFT	U		U								
Downloaded file: check	U		U								
Downloaded file: debit card	U		U								
Downloaded file: credit card	U		U								
Money order											
Cashier's check											

NOTE: "U" signifies that the information usually appears in the record. "S" signifies that the information sometimes appears in the record. Service dates refers to the dates that the service was received, such as the span of time that a utility provided service. "Item verbatim description" refers to text used by the vendor such as "tshrt" while "item description" refers to readily understandable text such as "tee shirt." "Item ID number" refers to a code such a SKU or UPC number. "Item price" refers to the price of individual items included in the "total price." "Item rebate price" refers to discounts for individual items.

Check Purchase with No Receipt. Respondent B signs a contract to have an addition built for his house. Respondent B gives the builder a $2,000 check but builder does not immediately provide a receipt for payment; the only record of the payment is the contract and the check itself. The checking account data file that Respondent B sends to the repository shows the date and amount of the check. No receipt matches this check. For that reason, the system asks Respondent B about that check in the web questionnaire: "On July 9, 2011 you wrote a check for $2,000. What did you purchase?" After Respondent B enters "Construction on my house," the system consults the database of terms, but does not find that phrase. The system therefore asks, "Was this check for food and drinks away from home, food and drinks for home consumption, clothing, shoes, jewelry and accessories, or all other products, services, and expenses?" Respondent B replies that the expense is in the "other" category. The web survey then asks the question, "Is this purchased for someone outside the household?" and Respondent B replies with "no."

Cash Purchase with Receipt. Respondent C buys $20.00 of gasoline and pays with cash. Respondent E gets a receipt for this purchase, and later scans the receipt and emails it to the repository. The system applies OCR to read the receipt, and tries to match the purchase with any check or credit card purchase in the bank and credit card files. Since this was a cash purchase, the system finds no match. It therefore asks Respondent C about the receipt in the web questionnaire. The software could be able to apply OCR to determine that the receipt is for gasoline, which falls into the "other" CE category. The software could also be able to capture the price from the scan of the receipt. The web questionnaire displays an image of the receipt and asks "This appears to be a receipt for a $20.00 purchase of gasoline. Is that correct?" When Respondent C answers "yes," the web questionnaire asks the question appropriate for the "other" category—whether the item was for someone on Respondent E's household list.

Electronic Fund Transfer with No Receipt. Respondent D pays her electric bill through electronic fund transfer from her checking account. No receipt is issued. The system detects this debit in the checking account data file and finds no matching receipt. Accordingly, the web survey asks, "On July 12, 2011, your checking account was debited $125.60 to the Southern Electric Company. What did you purchase?" It then asks the other questions necessary for the CE program.

Respondents Who Do Not Elect to Submit Their Data Electronically. Diary survey respondents who choose not to submit their data electronically would still be encouraged to keep their receipts. They would be given a box or envelope which they would use to store all their receipts. Printed on the box or envelope would be a reminder to obtain and keep receipts. The

envelopes or boxes would be ready for mailing, with a postage-prepaid frank, and pre-addressed to the data repository.

These respondents would also be asked to keep all their credit card, checking account, and other relevant statements that they receive in the mail. If they make any online purchases, they will be encouraged to print out and retain the purchase confirmations.

The respondents would be encouraged to create their own record when they make any purchases that involve no receipt. They would be provided short forms to record purchases. If they make a purchase with no receipt, such as from a vending machine, they would complete the form to create a record.

The respondents would periodically send, via U.S. Mail, their receipts and forms to the central repository. They would also mail photocopies of their credit card and checking account statements. Repository staff would scan the receipts and forms (much like services like shoeboxed.com do), and enter data from the financial statements.

Respondents Who Wish to Use a Paper Diary. Some respondents will not wish to scan or to mail their receipts to the central repository. These respondents will be permitted to record their expenditures in a paper diary, using the current Diary survey methods.

Interviewer's Role. During the initial visit to the household in the Diary survey, the interviewer would determine whether the respondent was willing and able to use technology to report their expenditure data. The interviewer would give instructions on how to record information. The nature of those instructions would depend upon the respondent's willingness and ability to use technology. Technology users would hear how to scan receipts and download data files. Nonusers of technology would hear how to keep and mail paper records. The interviewer would go through several examples with the respondent until the interviewer felt certain that the respondent understood.

The interviewer would set up advance appointments to talk with the respondent by telephone. The purpose of these calls would be to monitor and assist the respondent during the 2-week collection period. The interviewers would speak with the respondents at least four additional occasions after the first contact. The first call would occur 2 or 3 days after the initial contact.

The interviewer would monitor the activity of those respondents. The interviewers would observe if respondents were complying with the instruction to submit data to the repository, whether by U.S.

Mail or electronically. During the first call, interviewers who detect no data collection activity would ask the respondents whether they are obtaining receipts. If they are, the interviewers will encourage them to scan the receipts or send them in by U.S. Mail. If the respondents are neglecting to complete the web questionnaire, the interviewers would encourage them to do so.

The third contact would occur 1 week after the initial interview. The interviewers would have the option to make either an in-person visit or a telephone call. An in-person visit would be necessary if the respondent was failing to scan receipts or to send or mail the files to the central repository. In this case, the interviewer would personally examine the receipts and other documentation and conduct a recall interview. An in-person visit would also be necessary for those respondents who are reporting their expenditure data by filling out the paper diary.

The fourth contact would be by telephone. The interviewer would monitor the respondents' data collection activity. For respondents who are not sending information to the repository, the interviewer would ask about purchases that may have been missed.

The fifth, and final, contact would close out the data collection. If the respondent has been providing data electronically and completing the web questionnaires, then the contact could be by telephone, with a final review of purchases over the prior week. If the respondent had been providing data in paper form, or had elected to scan or mail in receipts but had not been consistently complying, this visit would be in person. The interviewer would conduct a recall interview during this visit.

Recall Interview. The recall interview would augment the data that the respondents had provided either electronically or on paper. The interview would also help the respondents to report expenditures that they had not yet reported. Some respondents will not adequately download financial data files or keep receipts. For those respondents, the recall interview will be the primary means of collecting data.

The respondents would provide details about their expenditures in the recall interview. The recall interview would be similar to the current Quarterly Interview survey, except that we recommend using an Event History Calendar to help the respondents recall their expenditures. An Event History Calendar uses salient events during the reporting period to help the respondent remember events that may not be as salient—such as expenditures.

Single or Multiple Respondents per CU. Considering the results of prior research (Westat, 2005; Edgar et al., 2006), we recommend a multiple respondent approach for the Diary survey in which all adults in the household provide data about their expenditures. The principal respondent would be solely responsible for reporting key expenditures such as rent and utilities. For those that report their data electronically, all CU members at least 14 years old will have their own account in the central repository. They would be asked to download and email their credit card, checking, and other relevant financial files. They will be asked to scan and email receipts. They will also be asked to respond to a web survey about their purchases. For those CUs that do not report their data electronically, all CU members will mail in receipts and financial statements or keep paper-and-pencil diaries. Following the procedures of FoodAPS and other surveys using multiple respondents (Westat, 2005; Edgar et al., 2006), the interviewer would ask the principal respondent to instruct the other household respondents in the data collection procedures.

Respondent Incentives. We suggest both monetary and non-monetary incentives. Monetary incentives have proven very effective on field studies, such as the Diary survey, to increase response rates in a cost effective manner.

Given the level of effort required for the CE Diary survey, we recommend an incentive of at least $50 for the principal respondent. This amount is slightly higher than the amount that Goldenberg et al. (2009) tested. However, it is consistent with the amounts proposed for the Medical Expenditure Panel Survey and several other field surveys such as the National Survey on Drug Use and Health. The incentive would be provided in the form of a prepaid debit card. We also propose a smaller incentive of $10 for each of the individual respondents within the household. This is consistent with several other federal surveys with multiple respondents (e.g., National Health and Nutrition Examination Survey, and FoodAPS). A final level of monetary incentives should be determined during the development process. FoodAPS will soon experiment with several incentive levels; the results could be applicable to the CE Diary survey.

Additional incentives might be useful to encourage respondents to use electronic records and to keep receipts. Wine and Riccobono (2011) report the use of incentives to promote early response to the web survey for the National Postsecondary Student Aid Study. A similar incentive structure could be considered for respondents who email and scan expenditure documentation in a timely way.

We also recommend that the household members receive a non-monetary incentive. The CE program provides an opportunity to offer unique incentives that are relevant to the survey itself.

That is, CE respondents might receive an incentive in the form of useful information about their spending. This incentive might motivate respondents to provide complete and accurate data because those data would actually enhance the value of the incentive.

One potential non-monetary incentive is a report in which the respondent's spending is compared with that of other people. The report would contain both graphs and text. This incentive would show respondents how their data compared with others with similar demographic background. For example, the incentive could be in the form of reports which contained graphs displaying the respondents' spending patterns beside the spending patterns of people of the same age and sex, of people who live in the same geographic area, or who have the same number of people in the household. The comparison data for these reports could come from recent CE results. These reports could be in hard copy or online form.

The financial software package Bundle uses a similar approach (www.bundle.com). Developed by Microsoft, Morningstar, and Citigroup, Bundle helps users monitor their spending and keep to a budget, like other financial software products such as Mint, Moneydance, and Buxfer. Bundle is unique, however, in that it creates online charts for the users, showing the characteristics of their spending alongside the spending of others with similar demographic backgrounds.

Smartphones. Some respondents will be interested in using their smartphones to keep some expenditure records. These respondents might use their smartphones to photograph purchases that have no receipts, such as an item bought from a vending machine. They might use their smartphones to fill out a short form, appearing on the screen, to provide data about the purchase. They might record an audio memo describing the purchase on the smartphone.

Respondents potentially could then refer to these reminders later, to complete paper forms about the purchases. They would then scan the forms and send them electronically to the repository.

Conceivably, the photographs, audio memos, or electronic forms could be sent electronically to the repository. These reminders could then be incorporated into the web survey. For example, one page on the web survey might have a question like this: "On July 20, 2011 at 3:19 pm you recorded this audio memo [click here to listen]. What did you purchase?"

We believe that smartphones potentially can play an important role in the CE program, as we described in our earlier report for BLS (Schneider and Reifer, 2011).

Loyalty Card and Other Retailer Information. For both the Diary and Interview surveys in our proposed CE redesign, respondents would be asked to sign a consent form to allow the interviewers to obtain expenditure information from the stores where the respondents have loyalty cards. Prior arrangements would have to be made with the retailer to arrange for release of this information. BLS, the Census Bureau, or a contractor acting as a government proxy, could attempt to negotiate procedures for obtaining data from grocery chains, chain pharmacies, and other retailers who have loyalty card programs.

The loyalty card data could serve several purposes:

- The data would reveal expenditures that the respondents neglected to report;
- The data could be used to check the accuracy of the data that the respondents provided from memory; and
- The data could be used to itemize purchases when the respondents provided only the total sum of all items purchased. For example, when respondents provide a credit card invoice that shows only the total amount of the purchase, these data could provide a list of the purchased items and their individual costs.

Diary Survey Costs and Implications for Sample Size

In this section we discuss the estimated costs and variance implications of the proposed procedures.

Costs. The Census Bureau provided a summary of CE variable costs for 2010 (including a breakdown between interview and diary costs). We used that summary as the primary basis for estimating the cost of the proposed design. We transformed some of the Census data (e.g., costs per interview, costs per noninterview) into more commonly used metrics (e.g., cost per completed interview as the expression of the total costs divided by the number of completed interviews). Because the current design does not incorporate any administrative data, we drew upon Westat's 2009 cost experience collecting data from administrative records of health care providers on the Medical Expenditure Panel Survey's Medical Provider Component (MPC), sponsored by the HHS Agency for Healthcare Research and Quality (AHRQ).

Table 9 summarizes the results of the cost estimation exercise. It compares the current design with the design we propose here, holding sample size constant, at 7,000 Consumer Expenditure (CE) units. This table provides costs for both the proposed Diary and Quarterly Interview survey design.

In this section we concentrate on the costs of the Diary survey. The Quarterly Interview survey costs are discussed in a later section.

Many assumptions are required to extrapolate costs for the proposed design from the Census CE data and the Westat MPC data. For collecting the Diary survey data, the assumptions are provided in Tables 10 and 11 for the field data collection and administrative record collection, respectively. Each table compares the proposed design with the current design. For the field data collection, these costs consist of a combination of personal visits and prompting by telephone. For the administrative data, these consist of contacting providers and collecting record data.

Table 9. Summary comparison of current and proposed CE designs

	Current design	Proposed designs	
		One respondent per CE unit for Quarterly Interview, multiple for diary	Multiple respondents per CE unit for Quarterly Interview, multiple for diary
Quarterly Interview			
Interviews/CE unit/quarter	1	1	2
Quarters	4	4	4
Interview cost	$17,048,783	$24,685,164	$43,199,036
Interview cost/complete	$476	$342	$375
Administrative records cost	NA	$10,192,920	$13,250,796
Administrative records cost/complete	NA	$283	$368
TOTAL QUARTERLY INTERVIEW COST	$17,048,783	$34,878,084	$56,449,832
TOTAL QUARTERLY INTERVIEW COST/COMPLETE	$118	$171	$150
TOTAL QUARTERLY INTERVIEW COST/CE Unit	$588	$969	$1,568
Diary			
CE Units	7,449	7,449	7,449
Diaries/CE unit/week	1	2	2
Weeks	2	2	2
Interview cost	$5,111,926	$7,502,577	$7,502,577
Interview cost/complete	$343	$252	$252
Administrative records cost	NA	$805,992	$805,992
Administrative records cost/complete	NA	803	$803
TOTAL DIARY COST	$5,111,926	$8,308,569	$8,308,569
TOTAL DIARY COST/COMPLETE	$343	$279	$279
TOTAL DIARY COST/CE UNIT	$686	1,115	$1,115
GRAND TOTAL – CE COST FOR ONE YEAR, BOTH COMPONENTS	**$22,160,709**	**$43,186,652**	**$64,758,401**

Table 10. Summary of CE diary costs

	A	B	C	D	E	F	G
	Total diary cost	Completed CE units	Cost/ completed CE unit	Completed diaries	Cost/ completed diary	Administrative records cost	Admin. records cost/ CE unit
Formula			A/B		A/D		F/B
Current design (Assumes 2PVs, one diary per week, 2 weeks)						NA	NA
PV	$5,111,926	7,449	$686	7,449	$686	NA	NA
TOTAL DIARY COST	$5,111,926	7,449	$686	7,449	$686	NA	NA
Proposed design							
(Assumes 2.3 PVs, Two diaries per week, 2 weeks)							
Initial visit (PV)	$4,355,375	7,449	$585	14,898	$585	NA	NA
3 Prompts,most by phone	$931,125	7,449	$125	14,898	$100	NA	NA
Follow-up Interview							
- Telephone @ .3	$391,073	2,235	$175	4,469	$215	NA	NA
- Personal Visit@.7	$1,825,005	5,214	$350	10,429	$400	NA	NA
TOTAL DIARY COST	$7,502,577	7,449	$1,007	14,898	$504	NA	NA
TOTAL DIARY + ADMIN. RECORDS	$8,308,569	7,449	$1,115	29,796	$279	$805,992	$108

Table 11. Administrative records cost for diary

	2009 MEPS MPC		Diary Proposed CE design		
	Hospitals	Office-based doctors	Large organizations	Small organizations	Sum
Interviewing	$2,339,952	$1,172,293			
Abstraction	$1,710,737	$631,824			
Other	$957,298	$510,209			
Total	$5,007,987	$2,314,326	$552,866	$253,126	$805,992
Completes (providers/vendors)	4,720	8,501	1,042	931	1,973
Cost per unit	$1,061	$272	$531	$272	$803
Household reporting units	8,500	8,500	7,444	7,444	
Provider or vendor completes/household reporting units	0.56	1.00	0.56	0.25	

The cost of the proposed design for the Diary is approximately 60 percent higher than the current design. One major addition for the new procedure is the collection of data from multiple respondents. The additional personal visits and telephone calls shown on the budget provide the additional costs associated with this. This feature contributes approximately $2.4 million of the additional $3.2 million of the higher costs of the new procedure. A second addition is the collection of administrative data, which is not done under the current design. This accounts for the remainder of the additional costs.

The administrative records aspect of the proposed design poses the greatest cost uncertainty. For this exercise, we have assumed that vendors and financial institutions will cooperate in providing cost information for their customers, but the feasibility of the proposed approach needs to be established with a proof of concept exercise, then tested carefully in the field. The experience base is much smaller for collecting administrative data in surveys than for conducting interviews and collecting diaries from household populations. We have made some rather arbitrary assumptions about the number of large and small vendors that would be asked to provide cost information for their customers, leveraging off the MPC data for hospitals and physicians. However, it should be noted that the MPC sample was driven by cost considerations. It supplements the MEPS Household data, and AHRQ determined from the outset that some provider types were more important than others for estimating national healthcare costs. Thus, hospital events reported by the household were sampled with certainty, but only a subsample of the office-based physicians was included in the MPC. In imputing for missing data, households with complete data for physicians served as donors for households with physician visits that were not selected for the MPC. Similar decisions could be made about the CE administrative data in the proposed design, so that it could be scaled up or down to meet available resources.

Implications for Sample Size. From a sampling and variance perspective, the proposed design does not significantly differ from the current design. Both use a 2-week data collection period. Therefore, the above cost discussion provides guidance on the sample sizes required to meet current program standards. As discussed, the cost of the recommended design, using multiple respondents, is approximately 80 percent more expensive than the current design. It is difficult to translate this exactly to implications for sample size. One approximation assumes a direct relationship between the increase in cost and sample size. Under this assumption, and without increasing the CE budget, the recommended design would collect data from approximately 60 percent of the CUs that provide data under the current design. Much of the additional collection cost for the new procedure is attributable to the collection of diaries from multiple respondents, rather than just one respondent.

If the current single respondent design were continued, the unit cost of the recommended design would be closer to the current design, even including the administrative record collection.

Recommended Design: Quarterly Interview Survey 5

In our recommended redesign, the primary purpose of the Quarterly Interview survey is to collect information about larger expenditures, while the primary purpose of the Diary survey is to collect data on smaller, more frequent expenditures. For example, expenditures for food would be covered by the Diary survey; these usually small, frequent expenditures would either not be covered, or be covered in only abbreviated form, by the Quarterly Interview survey. We are asumming that the income data that is currently collected on the Quarterly Interview survey would also be collected.

Our recommendations for the Quarterly Interview survey are intended to minimize respondent burden in two ways:

- The number of in-person interviews per CU will decrease from the current five to three; and
- Some expenditure data will be collected through records, rather than interviews.

The current design has a total of five in-person interviews per household, creating significant cost and respondent burden. Reducing this number to three in-person interviews would substantially reduce this burden and may lead to greater cooperation, fewer dropouts, and better data quality. The rotation design is illustrated in Figure 3. Each sampled address will be interviewed in two sets of interviews, spaced 12 months apart. For example, in Figure 3, the sample in panel 3 month 3 is initially interviewed in Year T+1 and then re-interviewed in Year T+2 in month 3. By collecting data at two points in time, the CE program can continue to collect longitudinal data to measure annual change and take advantage of the economies associated with following up households that have already been recruited.

The second way our recommended redesign reduces respondent burden is by collecting data about many of the expenditures through records. Data about most or all of the large expenditures is likely to be available on paper or electronic records and receipts. In addition, data about some expenditures may be available relatively easily from sources other than the household. For example, financial companies may provide data about mortgage and other loan payments. Utility companies may provide data about electricity or natural gas payments.

Figure 3. Rotation schedule for proposed design for the Quarterly Interview survey

Year T			Year T+1												Year T+2												Year T+3		
Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
																											5(1)	5(2)	5(3)
															4(1)	4(2)	4(3)	4(4)	4(5)	4(6)	4(7)	4(8)	4(9)	4(10)	4(11)	4(12)	4(1)	4(2)	4(3)
			3(1)	3(2)	3(3)	3(4)	3(5)	3(6)	3(7)	3(8)	3(9)	3(10)	3(11)	3(12)	3(1)	3(2)	3(3)	3(4)	3(5)	3(6)	3(7)	3(8)	3(9)	3(10)	3(11)	3(12)			
2(10)	2(11)	2(12)	2(1)	2(2)	2(3)	2(4)	2(5)	2(6)	2(7)	2(8)	2(9)	2(10)	2(11)	2(12)															
1(10)	1(11)	1(12)																											

Key:
X(Y) = Panel X interviewed in month Y
Example: 2(10) = Panel 2 is interviewed in October of Year T and in Year T+1

Data might also be collected for each respondent from retailers who have loyalty card programs. However, these expenditure data would not be collected for food expenditures in the Quarterly Interview survey; food expenditure data would be collected only in the Diary survey.

By increasing the focus on records, the goal is to reduce the overall difficulty of the data collection for the individual respondent. In addition, it is to reduce the length of the interview. The (untested) assumption is that with greater use of records, the overall interview would be shorter.

Contacts and Interviewing. Table 12 summarizes the data collection. In the first year, there would be three telephone contacts between two in-person contacts. The first in-person contact initiates the procedures. The respondent will be asked for consents to release records, given instructions on the procedures for record collection, and participate in a bounding interview. Two weeks, 1 month, and 2 months after the contact will be telephone calls to check on the respondent's progress and to encourage the respondents to continue providing data. During the 3-month in-person contact, the interviewer will review all records and documentation that have been collected and administer a recall interview for the 3-month period.

Twelve months after the initial contact, a packet will be mailed to the CU to re-orient the respondent to the data collection tasks. The interviewer will follow this up with a telephone call to confirm the procedures and conduct a bounding interview. Two weeks later, 1 month later (month 13), and 2 months later (month 14), the respondents will receive reminder phone calls. At month 15, an in-person contact will occur using the same procedures as the 3-month in-person contact.

First in-Person Interview. The first interview with the household would be completed as a personal visit. During this visit, the interviewer would identify a respondent in the CU who was willing to report on all expenditures for everyone in the household. The respondents would tell the interviewers whether they were willing and able to download the data files for their checking, credit card, or other accounts.

The respondents would be asked to sign two consent forms. One would allow survey staff to obtain expenditure data directly from the respondents' utility and mortgage companies. These expenditure data would not be collected from the respondents who sign this consent. The second consent would allow survey staff to use the respondents' loyalty card numbers to obtain expenditure data from the retailers that issued the loyalty cards.

Table 12. Quarterly Interview survey schedule and procedures for sampled CUs

Initial interview - In-person interview
- Obtain consents
- Determine data collection methods, electronic or paper
- Provide data collection instructions
- Conduct bounding interview

2 weeks, Month 1 and Month 2 after initial interview – Telephone contact
- Review respondent's provision of data, electronic or paper
- Encourage respondent to provide data

3 months after initial interview – In-person contact
- Review records collected
- Conduct recall Interview

12 months after initial interview – Telephone/In person
- Mail instruction packet beforehand
- Followup Telephone call to orient Respondent and conduct bounding interview
- If household has changed, use initial contact protocol

12 months plus 2 weeks, 13 months, 14 months after initial interview – Telephone contact
- Review respondent's provision of data, electronic or paper
- Encourage respondent to provide data

15 months after initial interview – In-person interview
- Review records collected
- Conduct recall Interview

The interviewer would then administer the Quarterly Interview survey, to collect information on recent purchases. The last month might be a suitable reference period. In this way, the first interview will serve as a bounding interview for the second interview.

The interviewers would instruct the respondents on the data collection procedures. All respondents would be instructed to save receipts for all their expenditures, except food and alcoholic beverages. They would also be instructed to write a note for all expenditures for which they had no receipt. They would receive short forms to use for those notes.

Additional instructions would vary, depending upon the individual respondent's ability and willingness to download data files for their checking, credit card, and other relevant accounts. Respondents who will download these files would receive instructions similar to the instructions given to the Diary survey respondents who will download these files. The procedures for the Diary survey are explained above but will be summarized here.

The respondents would be asked to download their credit card, checking, and other account data files and email them to the central repository at least once per week. They would also be asked to keep their receipts from all purchases except food. They would be given a small receipt scanner and software developed specially for the CE program. The respondents would scan their receipts and email the scan files to the central repository at least once per week. They would also scan the notes that they created for purchases with no receipts, and email the scan files to the central repository. They would be given special software to automate all these procedures.

The CE system would derive the individual expenditure items from the receipts. It would use OCR software to turn the scanned receipts into text. Software created especially for the CE program would extract from this text the description and price of each individual expenditure item. The CE system would match the expenditure data in the scanned receipts with expenditure data from the downloaded credit card, checking, and other account files.

The respondents would periodically complete a web questionnaire. The web questionnaire would ask about each individual item purchased. The questionnaire would ask details needed by the CE program such as whether the expenditure was for someone outside the household.

Some respondents may be unwilling or unable to download financial files, or to scan receipts, or to complete a web survey. These respondents would be given envelopes to store their paper receipts and completed forms for purchases with no receipts. The envelopes will carry a postage prepaid frank, and be pre-addressed to the central repository. The respondents would be asked to mail an envelope on a regular schedule. CE staff would then receive the envelope and scan the receipts. The respondents would also be asked to mail in photocopies of their monthly statements for their credit card, checking, and other accounts. CE staff would enter those data as well.

Interim Contacts. Field representatives would monitor the activity of each respondent. They would determine whether each respondent was scanning receipts and forms and emailing the scan files to the central repository, or mailing envelopes with paper receipts and forms. The field representatives will call the respondents two weeks after the first interview to ask if the respondent had any questions or concerns. If the respondent has not been sending data to the central repository, or appears to be sending too little data, the field representative will determine whether the respondent is encountering any problems or is simply remiss. The field representative would try to address any issues that the respondent may be encountering.

The field representatives would telephone each respondent three times: 2 weeks, 1 month, and 2 months after the first interview. Again the field representatives would attempt to ensure that the respondents remained motivated to provide data for their households. They would ask the respondents about any potential problems. They would check the repository to determine which respondents had stopped providing data, or seemed to provide data only sporadically. The field representatives would try to remedy any issues during these telephone calls.

Second in-Person Interview. During the second in-person interview, conducted 3 months after the first in-person interview, the field interviewer will review the receipts and forms that the respondents sent to the repository, and, if the respondent used the Internet, the data that the respondent provided through the web questionnaires. The interviewer would ask the respondent to provide information that may have been omitted in the prior web questionnaires for any reason, such as the respondent's failing to know at the time about an expenditure by another member of the household, or the respondent's neglecting to obtain or scan a receipt. If the respondents did not complete any web questionnaires or had provided few receipts and data files, the interviewer would administer a recall interview.

The interviewer's questions would take into account the reference period for each kind of expenditure. The interviewers would ask about a 1-month, 3-month, or 12-month period, depending on the nature of the expenditure. The length of this interview would depend on how much information the respondent had sent to the central repository. For example, if the respondent had already sent in information for certain types of expenditures, such as regular mortgage payments, utility bills, health club memberships, health insurance, the interviewer would not need to ask questions about these expenditures. The interviewer would ask about expenditures which the respondent may not have exhaustively reported.

The interviewer would encourage the respondent to use records as a memory aid during this interview. For example, some respondents may have receipts on hand that they never sent to the central repository. Some respondents who did not download their financial files for their credit card, checking or other accounts may have hard-copy statements for those accounts.

Other memory aids should be developed as the structure of the interview is determined. An event history calendar could be used, at least for those respondents who have not provided much information to the central repository. All materials should be designed to facilitate the interview. The material shown to respondents should be optimized for memory search. For example, research using eye-tracking and response latencies (Redline et al., 2009) has found grouping responses within major

subheadings, such as on the current CE Information Cards, can lead to confusion. The formatting for the visual aids should be tested so that it can help respondents provide information to the interviewers. Novel types of displays should be tested, such as pictures of items, which might be easier to coordinate with interviewer questions.

Contacted at Month 12. This contact would initiate a 3-month period in which the respondent would report data. Prior to the interview, a field representative would mail the respondent a package that included the instructions and materials for the survey. If this is the same CU as the prior contact, a telephone call would be made by the field representative to re-acquaint the respondent with the survey procedures. The field interviewer would remind the respondent to provide electronic records. The field interviewer would attempt to elicit cooperation from respondents who were not providing data diligently. The interviewer would conduct a brief recall interview which would serve as a bounding interview.

An in-person interview would be conducted for households that have changed since the first interview. The in-person interview is required in that situation because the interviewer must obtain consents, provide instructions and conduct the bounding interview. Otherwise, this interview would be conducted by telephone.

Interim Contacts and Interview at Month 15. The methods for the interim contacts and the recall interview at month 15 would repeat the same procedures as conducted 12 months earlier.

Number of Respondents. We are aware of no studies that examined the feasibility of using multiple respondents in each household in the Quarterly Interview survey. Prior research, described above, suggests that the Diary survey, with its short time frame, could benefit from a multiple-respondent design. Further research is needed to determine whether a multiple-respondent design is cost-effective for the Quarterly Interview survey.

If the decision is made to stay with a single-respondent design, techniques could still be developed to improve data quality. Most importantly, respondents could be encouraged to speak with other household members daily to learn about their expenditures. Each household member could receive a box or an envelope for storing receipts or forms documenting purchases with no receipts.

Incentive. We recommend that respondent incentives in the Quarterly Interview survey be similar to those for the Diary survey. Both a monetary and a non-monetary incentive should be provided. The monetary incentive would be $50 for each of the two recall interviews with the CU respondent.

A smaller incentive, perhaps $5, would be given to other household members to provide receipts and other documentation of purchases. The non-monetary incentive would be a financial report on the expenses of the household as described above in the section on the Diary survey.

Monitoring Interviewers. The success of the Quarterly Interview survey depends in large measure on the care with which the interviewers collect the data. In particular, the interviewers need to probe for detailed information when necessary. Some respondents may tend to prematurely decide that they have no more data to report (Biderman et al., 1986). Appropriate interviewing methods, such as the use of extensive cueing can combat this tendency. Interviewers must have the skill to balance the need to probe for details and the need to avoid burdening the respondents.

Past research suggests that when supervisors monitor interviewers' activities, the interviewers are less likely to take shortcuts and more likely to attempt to elicit survey data thoroughly. Two field experiments involving the National Crime Victimization Survey (NCVS) suggested that centralized telephone interviewing can produce a higher number of crime victimization reports than decentralized field interviews can produce (Hubble and Wilder, 1988; Biderman et al., 1986). The reason for this effect may be that the telephone interviewers' data collection activities were monitored.

An important component of the CE redesign should be to institute methods to train, monitor and provide feedback to interviewers on an ongoing basis. Dijkstra et. al. (2008) discussed the importance of these aspects of the design when administering the SIPP Event History Calendar. The monitoring can be done through use of paradata from the computerized interview (e.g., section timings) and the use of Computer Audio Recorded Interviews (CARI). The Census Bureau is now instituting CARI within many of their household surveys. With CARI, entire interviews or just samples of interviews can be recorded. Supervisors later review these recordings to check interviewer practices and provide feedback on performance.

Costs of Proposed Design

For the Quarterly Interview component, Table 9 provides two variants, the "one respondent per CE unit" model like the current design, and an alternative that collects data directly from multiple people in the unit. Tables 13 and 14 provide the assumptions behind the field data collection. Table 15 provides the costs of collecting administrative records.

Table 13. Summary of Quarterly Interview cost per year

	A	B	C	D	E	F	G	H	I
	Total interview cost	Quarters of completed CE units	Cost/ completed CE unit	Number of completed interviews	Cost/ completed interview	Administrative records cost	Admin. records cost/ CE unit	Interview plus records cost	Interview plus records cost/ CE unit
			A/B		A/D		F/B	A+F	H/B
Current design									
One respondent per CE unit	$17,048,783	29,000	$588	35,843	$476	NA	NA	$17,048,783	$588
Proposed design									
If one respondent per CE Unit	$24,685,164	36,000	$686	72,074	$342	$10,192,920	$283	$34,878,084	$969
If multiple responders per CE unit*	$43,199,036	36,000	$1,200	115,148	$375	$13,250,796	$368	$56,449,832	$1,568

*Assume interview costs are twice the cost of one respondent per CE unit, admin. records costs are 1.3 times the cost of one respondent per CE unit.

Table 14. Quarterly Interview costs by mode (assume one respondent per CE unit)

	Current design					Panel 1 Proposed Quarterly Interview design				Proposed Quarterly Interview design Full Implementation, all panels Steady state (Data for 7,000 CE Units/Quarter)			
				$/Interview									
	Cost, all quarters	Cost per quarter	N	PV only	T only	Cost, all quarters	Cost per quarter	N	$/Int G/H	Cost, all quarters	Cost per quarter	N	$/Int K/L
	$17,048,783					$12,009,859				$24,685,164			
Quarter 1		$3,892,024	7,316	$532			$7,850,794	7,316	$1,073		$6,171,291	7,316	$844
Detail: Bounding Interview at 0 weeks (PV)		$3,892,024	7,316	$532			$3,892,024	7,316	$532		$2,479,030	7,316	$339
Prompt 1 at 2 wks (T)							$177,673	7,316	$24		$88,836	3,658	$24
Prompt 2 at 4 wks (T)							$177,673	7,316	$ 24		$88,836	3,658	$24
Prompt 3 at 8 wks (T)							$177,673	7,316	$24		$88,836	3,658	$24
Recall at 13 weeks (PV/T currently, PV proposed)		$3,417,737	7,092	$298	$146		$3,425,751	7,092	$483		$3,425,751	7,092	$418
Quarter 2 at 26 weeks (PV/T)		$3,276,573	7,101	$298	$146						$6,171,291	7,316	$844
											$2,479,030	7,316	$339
											$88,836	3,658	$24
											$88,836	3,658	$24
											$88,836	3,658	$24
											$3,425,751	7,092	$418
Quarter 3 at 39 weeks (PV/T)		$3,231,239	7,221	$298	$146		NA	NA	NA		$6,171,291	7,316	$844
											$2,479,030	7,316	$339
											$ 88,836	3,658	$24
											$88,836	3,658	$24
											$88,836	3,658	$24
											$3,425,751	7,092	$418

Table 14. Quarterly Interview costs by mode (assume one respondent per CE unit) (continued)

	Current design					Panel 1 Proposed Quarterly Interview design				Proposed Quarterly Interview design Full Implementation, all panels Steady state (Data for 7,000 CE Units/Quarter)			
				$/Interview									
	Cost, all quarters	Cost per quarter	N	PV only	T only	Cost, all quarters	Cost per quarter	N	$/Int G/H	Cost, all quarters	Cost per quarter	N	$/Int K/L
Quarter 4		$3,231,239	7,221	$298	$146		$4,159,065	7,221	$576		$6,171,291	7,316	$844
Recall Interview at 52 weeks (PV/T currently) @		$3,231,239	7,221	$298	$146						$2,479,030	7,316	$339
Bounding Interview at 39 weeks (T proposed)		$—					$841,755	5,777	$146		$88,836	3,658	$24
Bounding at 39 wks (PV proposed, new CE units)		$—					$768,297	1,444	$532		$88,836	3,658	$24
Recall Interview at 52 weeks (PV proposed) @							$2,549,013	7,221	$353		$88,836	3,658	$24
											$3,425,751	7,092	$418

Table 15. Administrative records cost for Quarterly Interview

	2009 MEPS MPC		Proposed CE design for Quarterly Interview		
	Hospitals	Office-based doctors	Large organizations	Small organizations	Sum
Interviewing	$2,339,952	$1,172,293			
Abstraction	$1,710,737	$631,824			
Other	$957,298	$510,209			
Total	$5,007,987	$2,314,326	$8,615,320	$1,577,600	$10,192,920
Completes (providers or vendors)	4,720	8,501	16,240	5,800	
Cost per provider or vendor	$1,061	$272	$531	$272	
HH Reporting Units	8,500	8,500	29,000	29,000	
Provider or Vendor Completes/Household Reporting Unit	0.56	1.00	0.56	0.20	

The total cost of the proposed design is approximately two times as expensive as the current design ($17 million vs. $34 million). Approximately $7 million of the increase is associated with the additional effort to collect the quarterly data from respondents. The proposed design requires three in-person contacts and a telephone contact to collect two quarters of data. To collect four quarters of data, the design then needs six personal visits and two telephone interviews. The current design uses 5 personal visits to collect the same number of collection quarters. In addition, two of the six in-person contacts for the proposed design are initial visits to the household, which are more expensive than follow-up visits. Only one of the five interviews for the current design is an initial visit. Finally, for the current design, there are interim contacts made by telephone during each collection quarter, which are not done for the current design.

The remaining additional cost of $10 million is from the administrative record collection. These costs could be cut considerably if the collection was restricted to utilities (e.g., gas, water, electricity). This is discussed further in the last section of this chapter when considering variations in the proposed designs.

In addition to the caveats described for the Diary costs, several others should be mentioned for the Quarterly Interview survey. The current design for the Quarterly Interview is multi-mode. Almost all of the Quarter 1 interviews are conducted by means of a personal visit or a visit in combination with some telephone interviewing. But in subsequent quarters many cases are completed through telephone interviews only, or through a combination of phone and personal visit. Therefore, it is very difficult to determine current design costs by mode. In the proposed design, we attempt to avoid this problem by assuming that all CE units provide data either entirely in person (as in Quarter 1) or entirely by telephone or by mail (as in the proposed Quarter 4 data collection). Census cost data for interviews conducted only in personal visits and for interviews conducted only by telephone are used to estimate costs for the proposed design.

Sample Size and Precision

Assessing the implications of the above costs for the sample sizes of the Quarterly Interview survey is not as straightforward as for the Diary survey. Whereas the sample design for the Diary is very similar between the proposed and current surveys, the equivalent Interview surveys are not. Consequently, to evaluate the implications of the above recommendations an analysis was conducted that simulated the proposed Quarterly Interview survey design. It was assumed that the survey

would produce annual estimates for the categories published by BLS (see http://www.bls.gov/cex/2009/stnderror/age.pdf). The expenditures used in the simulation were selected based on current practice. The decision on whether to use the Interview or Diary survey for estimates is based on the proportion of CUs that report the expenditure and the reported dollar amounts of the expenditure. The expenditure item has to be reported frequently enough on the survey to yield a reliable estimate. If the item is not purchased very often, then the Diary is not likely to be very reliable given the short reference period. If respondents tend to forget the expenditure, the Quarterly Interview survey may not be reliable. When differences in reliability are not as much of an issue, the decision on which survey to use is based on the assumption that the source yielding the higher expenditure amount is preferred.

For example, virtually all of the food and alcohol expenditure estimates are based on data from the Diary survey (see Appendix A for assignments in 2009). All estimates for housing expenditures are based on data from the Quarterly Interview survey. However, for a number of domains, the estimates for some expenditures use data from the Quarterly Interview survey and the estimates for others use data from the Diary survey. For example, estimates for housewares expenditures are based on data from both the Quarterly Interview (e.g., plastic dinnerware, flatware) and the Diary survey (e.g., china and glassware).

The expenditures used to simulate precision are shown in Table 16. These are the lowest level published categories that are exclusively or primarily taken from the current Quarterly Interview survey. The analysis used a public use version of the CES 2009 data set, which included interview data during the four quarters of 2009 and the first quarter of 2010. The analysis compared the estimated variance under the current design, $\hat{\sigma}^2_{tot}$, to estimates of variance for the alternative design, simulated using the same data set (see Appendix B Table 2 for the coding of the expenditure categories).

The first step was to identify the contribution of between-PSU variance to the overall variance. The scope of the research concerns the effect of redesigning the sample within the sampled PSUs, while holding the random selection of PSUs fixed.[2] Because the CES is essentially an unclustered sample, the specification to for variance estimation was changed using software (SUDAAN) to treat each consumer unit as a cluster, but to assume a simple random sample of consumer units had been selected. The calculation recognized that consumer units could contribute up to the annual results up to four times. Because the calculation involved the many thousands of observed consumer units

[2] See Appendix B for more details on the methods used for the simulation.

rather than the 44 replicates, the calculation of within variance is much less subject to random error. Because the same data and weights were used, the estimates of expenditures were identical to the first calculation but reduced estimates of the standard errors, $\hat{\sigma}^2_{wc}$, for all but 2 of the 53 estimates. The between variance, $\hat{\sigma}^2_b$, was estimated from the familiar decomposition for two-stage samples,

$$\sigma^2_{tot} = \sigma^2_b + \sigma^2_{wc}$$

and substituted the estimated values for $\hat{\sigma}^2_{tot}$ and $\hat{\sigma}^2_{wc}$ into this equation, setting $\hat{\sigma}^2_b = 0$ in the two exceptional cases.

Table 16. Expenditure types used to simulate proposed design for the Quarterly Interview survey by lowest level publication group

Expenditure type	Expenditure type
Mortgage interest and charges	Cars and trucks, used
Property taxes	Other vehicles
Maintenance, repairs, insurance, other expenses	Gasoline and motor oil
Rented dwellings	Vehicle finance charges
Other lodging	Vehicle, Maintenance and repairs
Natural gas	Vehicle insurance
Electricity	Vehicle rental, leases, licenses, and other charges
Fuel oil and other fuels	Airline fares
Telephone services	Public transportation except airfares
Water and other public services	Health insurance
Personal services	Medical services
Other household expenses	Prescription drugs
Household textiles	Medical supplies
Furniture	*Fees and admissions*
Floor coverings	Audio and visual equipment and services
Major appliances	Pets, toys, hobbies, and playground equipment
Small appliances, miscellaneous housewares	Recreational vehicle purchases
Miscellaneous household equipment	Other entertainment supplies, equipment, and services
Apparel and services, Men, 16 and over	Personal care products and services
Apparel and services, Boys, 2 to 15	Reading
Apparel and services, Women, 16 and over	Tuition
Apparel and services, Girls, 2 to 15	Other education
Apparel and services, Children under 2	Tobacco products and smoking supplies
Footwear	Miscellaneous
Other apparel products and services	Cash contributions
Cars and trucks, new	Life and other personal insurance
	Pensions and Social Security

The first three columns of Table 17 provides the expenditure estimates, the estimated standard errors for the current design, $(\hat{\sigma}^2_{tot})^{1/2}$ and the percent of the total variance due to between-PSU variance for expenditures using a 3-month recall period. For most expenditures (e.g., vehicle insurance, footwear), this percentage is under 50 percent indicating that the within-PSU variance

contributes most of the variance, so that any changes in the survey design within PSUs can have significant effects on the total variance. There are a few items with significant between PSU variance, including electricity, natural gas, fuel oil and water. For these items, the proposed changes will have limited effects on the total variance.

Table 17. Simulation of effects on within PSU variation of the redesigned Quarterly Interview survey for 1-, 3-, and 12-month reference periods by expense categories

Expense category	2009 estimate	Standard error	% between PSU variation	% reduction in within PSU variance		
				1 month recall	3 month recall	12 month recall
Mortgage interest and charges	3593.71	75.0	28	54	58	56
Property taxes	1809.44	31.1	45	57	60	56
Maintenance, repairs, insurance, other expenses	1138.71	43.1	33	-67	19	56
Rented dwellings	2854.51	50.7	10	52	57	56
Other lodging	669.01	30.2	27	9	49	56
Natural gas	483.4	16.8	85	37	52	56
Electricity	1376.65	29.2	89	44	53	56
Fuel oil and other fuels	141.46	9.2	68	-28	43	56
Telephone services	1161.76	12.1	56	48	54	56
Water and other public services	480.79	10.5	76	0	51	56
Personal services	389.44	20.2	0	53	45	56
Other household expenses	616.34	14.1	41	10	44	56
Household textiles	84.55	3.2	31	-141	14	56
Furniture	341.22	14.5	12	-191	14	56
Floor coverings	30.31	3.3	25	-241	1	56
Major appliances	188.77	8.8	29	-180	4	56
Small appliances, miscellaneous housewares	58.42	2.7	24	-427	10	56
Miscellaneous household equipment	461.1	14.8	40	-99	16	56
Apparel and services, Men, 16 and over	199.7	8.7	40	4	16	56
Apparel and services, Boys, 2 to 15	61.11	2.1	17	-100	34	56
Apparel and services, Women, 16 and over	336.41	8.8	47	-86	37	56
Apparel and services, Girls, 2 to 15	77.86	2.8	23	-114	37	56
Apparel and services, Children under 2	70.86	2.8	36	-25	44	56
Footwear	147.32	4.3	59	-116	23	56
Other apparel products and services	229.01	11.6	5	-249	29	56
Cars and trucks, new	1297.15	83.8	19	-275	0	56
Cars and trucks, used	1304.35	70.6	23	-131	6	56
Other vehicles	55.45	11.7	11	-221	3	56
Gasoline and motor oil	1986.4	20.6	40	44	50	56
Vehicle finance charges	281.01	8.2	35	59	62	56
Vehicle, maintenance and repairs	682.29	15.6	41	-135	20	56
Vehicle insurance	883.86	20.8	76	-138	38	56
Vehicle rental, leases, licenses, and other charges	446.12	16.0	42	18	53	56
Airline fares	300.92	11.2	18	-130	21	56

Table 17. Simulation of effects on within PSU variation of the redesigned Quarterly Interview survey for 1-, 3-, and 12-month reference periods by expense categories (continued)

Expense category	2009 estimate	Standard error	% between PSU variation	% reduction in within PSU variance		
				1 month recall	3 month recall	12 month recall
Public transportation except airfares	169.14	7.7	34	-80	22	56
Health insurance	1784.3	26.9	35	41	53	56
Medical services	739.89	25.7	49	-71	26	56
Prescription drugs	361.72	10.6	45	29	44	56
Medical supplies	86.9	4.8	19	-317	5	56
Fees and admissions	602.05	18.5	47	6	42	56
Audio and visual equipment and services	955.3	14.7	41	-35	20	56
Pets, toys, hobbies, and playground equipment	442.18	13.4	31	-72	45	56
Recreational vehicle purchases	160.1	30.2	10	-357	1	56
Other entertainment supplies, equipment, and services	209.53	9.6	33	-78	9	56
Personal care products and services	300.71	6.2	55	44	50	56
Reading	108.99	2.9	51	38	44	56
Tuition	886.87	40.0	0	-101	23	56
Other education	131.85	5.5	21	-72	24	56
Tobacco products and smoking supplies	377.64	10.6	23	47	48	56
Miscellaneous	747.07	31.4	2	-48	29	56
Cash contributions	1723.08	84.4	55	9	25	56
Life and other personal insurance	309.03	12.8	40	-139	37	56
Pensions and Social Security	5163.74	81.0	37	55	59	56

* Negative number indicates an increase in variance.

For all expenditure categories, the analysis simulated the effect of interviewing households for 1-, 3- and 12-month reference periods (see last three columns in Table 17). Examining the estimates for a 3 month reference period indicates the effects of eliminating the four-interview panel design, with no change in the reference period. To simulate the effect of interviewing households for one 3-month period, we changed the specification for the variance software to treat each observation in the 2009 data set as independent. This specification can be said to be unclustered. Again, the same estimates of expenditures for each group were obtained, but an estimated standard error, $\hat{\sigma}^2_{wu}$, smaller than $\hat{\sigma}^2_{wc}$. (In one case, the estimated percent reduction rounds to 0.) This represents the estimated reductions in variance from switching to the new design but using the current 3-month recall for all expenditures. We estimated the proportional reduction in within variance using

$$\hat{R}_1 = (\hat{\sigma}^2_{wc} - \hat{\sigma}^2_{wu}) / \hat{\sigma}^2_{wc}$$

These results are shown in the fifth column of Table 17. The within-PSU variance is significantly reduced in all cases. The median change across all expenditure categories is 37 percent.

We next considered the effect of changing the reference periods. This is provided in the fourth and sixth columns of Table 17. For a 12-month recall period it is possible to derive a theoretical expectation for the reduction in the within variance through a relatively simple argument. For any category in which 12-month recall is implemented, a single interview would provide the same information as available currently under four cumulative interviews. With respect to within variance, this change would multiply the effective sample size by a factor of 4. This simple argument implies a 75 percent reduction in the within variance.

The argument must be refined, however, because the assumed data for a given year would not be complete until the end of the fourth quarter of the following year. For example, with a 12-month reference period, respondents in 2011 can contribute to 2010 as late as November of year 2011. To prevent a change in schedule, the estimation must depend only on the first quarter of the following year. For the January of the current year, there will be reports from all 12 possible interview months up through January of the next year. Similarly, all 12 months are available for February and March, but April is reduced to 11, May to 10, up to only 3 for December. Thus the average of (1/12, 1/12, 1/12, 1/11, …, 1/4, 1/3), .14749, should be compared to 1/12, .08333. Thus, the variance, instead of .25 times the current variance, should be (.14749/.08333)*.25 = .4424, or a 56 percent reduction in variance instead of 75 percent.

In some cases, the estimated values for the reduction in the 3-month period exceed the theoretical value reported for 12-month recall groups, but never by a large amount. The estimates are affected by sampling error, but the large number of degrees of freedom used in the calculation provides a greater measure of stability than is frequently seen when analyzing components of sampling error.

Within the 12-month group there are distinct patterns according to the type of expense. Those expenditures that are done on a regular basis, such as utilities, health insurance, vehicle finance charges, there is practically no difference between the 3- and 12-month recall periods. This indicates that for these items, eliminating repeated interviewing during the calendar year provides most of the gains in precision. There is very little gain from extending the reference period to 12 months. As noted above, the design would target these types of expenditures for collection directly from utility companies, which makes the reference period less relevant as well. This pattern does not hold for items that are not purchased at high frequency (e.g., floor coverings; major appliances; vehicles) or

on a less regular basis (tuition; airline fares). In these cases, the 12-month period has a significant impact on the within-PSU variance.

The effect of a 1-month recall period was estimated using expenditures reported for the month preceding the interview. In this case, the estimated expenditures changed as well as the estimated standard errors. We again computed the unclustered variance for this design and estimated the (signed) reduction in variance by

$$\hat{R}_2 = (\hat{\sigma}^2_{wc} - \hat{\sigma}^2_{wu1}) / \hat{\sigma}^2_{wc}$$

Unlike the other estimates, there are significant increases in the variance, as indicated by the negative signs in the figure. Some of these items could have reference periods for much longer periods of time (e.g., auto sales).[3] There a number of items, however, that are negative and could not easily be moved to a 1-month or a 12-month period without significant loss in precision. On the other hand, there are items that have positive values, some of which are greater than 20 percent, some greater than 40 percent. These items should be considered when changing to a 1-month reference period.

We next translated the within-PSU reductions to reductions in total variance. These are shown in Figures 4–6, respectively, for each of the three different reference periods (see Appendix C for detailed data). Total variance reductions were calculated using the values of $\hat{R}_2$ and the values of the between variance. For example, even though the theoretical value for $\hat{R}_2$ of 56 percent is reported for 12-month recall groups, the effect on total variance is lessened when between-PSU variance is taken into account. For purposes of cost comparisons, we also calculated the effect of reducing the within-PSU sample sizes by 20 percent or 40 percent, not to propose these specific reductions, but to offer a wide range for consideration.

For a 3-month reference period, there are significant reductions in total variance across most expenditures. The median is 35 percent. Reducing the sample size by 20 percent brings variances close to the current design (median = 9%). Cutting as much as 40 results in about half of the expenditures increasing in variance and half still maintaining a reduction in variance (median = -4%).

[3] See Appendix C, Table C3, for the results for specific types of expenditures.

Figure 4. Percent reduction in sampling variance for a 3-month reference period

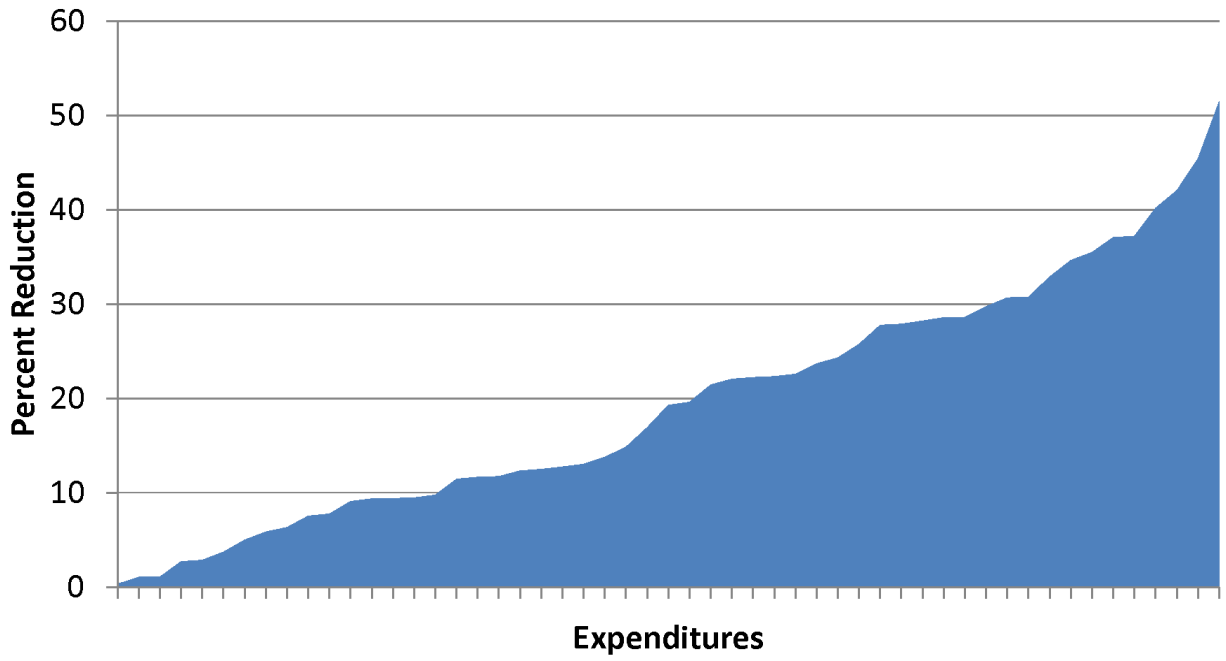


Figure 5. Percent reduction in sampling variance for a 12-month reference period

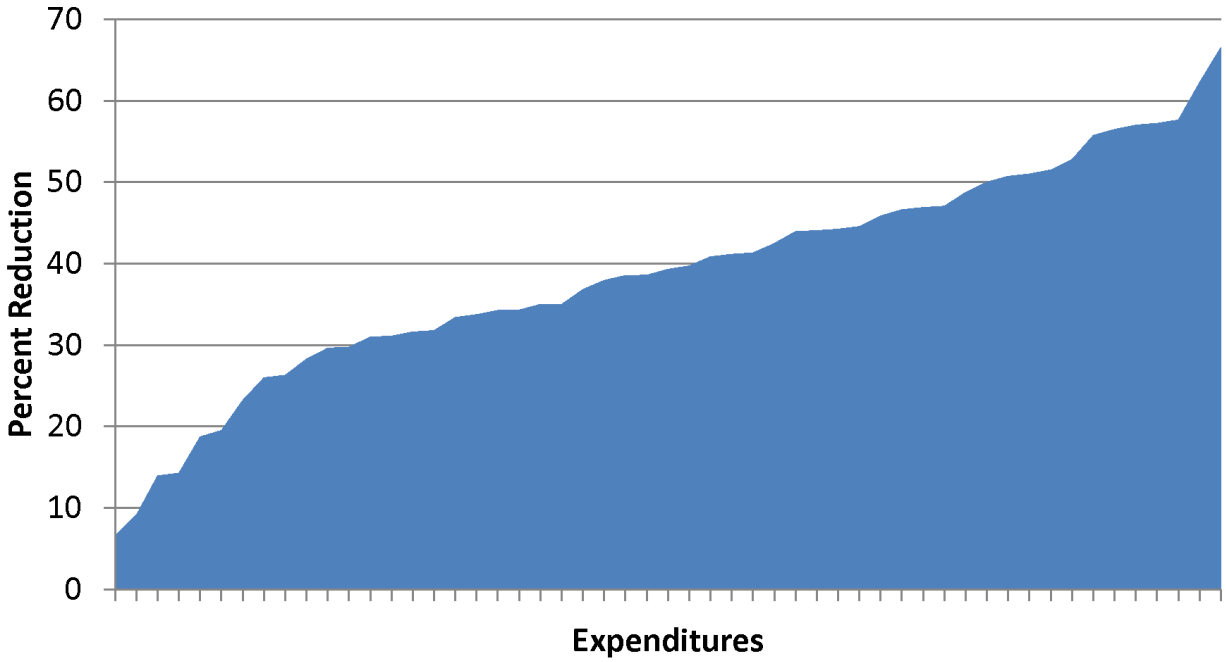
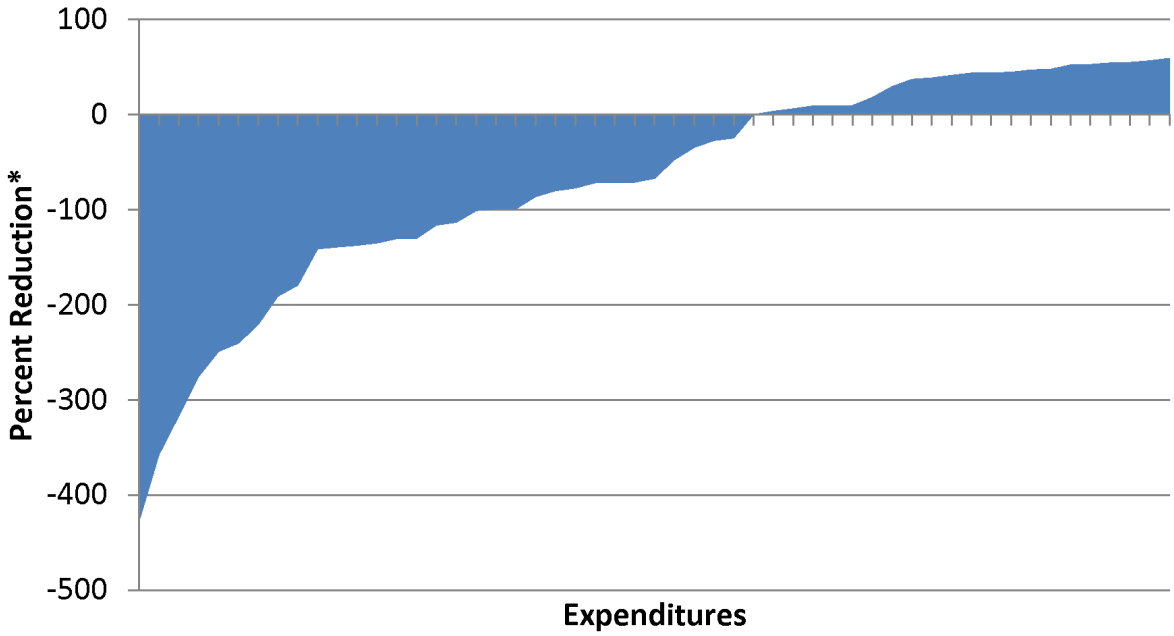


Figure 6. Percent reduction in sampling variance for 1-month reference period



* Negative number indicates in increase in sampling variance

As would be expected, switching to a 12-month period shows significant gains in precision, with a median of 36 percent with current sample sizes. Even with a 40 percent sample cut, most expense categories show positive reductions (median =18). The pattern is the opposite for a 1-month period. This results in significant losses in precision.

These results raise a question of whether a 12-month period provides net advantages, at least for the precision of the estimates. Switching to a 12-month period does not have a significant effect on variance for the regular expenditures examined in the present analysis (e.g., utilities). For these, almost all of the benefit of the proposed design is from altering the panel component of the current design. However, this does suggest that collection of these data only needs to occur for half of the interviews, if a 12-month period is to be used. For less frequent purchases, it does significantly increase precision. If the combination of the use of records and an enhanced recall interview yields accurate reports, a 12-month period could be a benefit. However, using a 3-month period would still have significant savings in terms of variance.

The above analysis could not simulate how the measurement would change when shifting to 1- or 12-month reference periods. For the 1-month simulation, the most recent month was used from a 3-month recall period. This reflects both possible omissions of expenditures, as well as telescoping.

Presumably asking for reports over a 1-month period would reduce both of these types of errors. Similarly, asking for recall over a 12-month period would likely increase measurement error, all other things being equal. Enhanced use of records and recall procedures would seek to improve this measurement.

Discussion

The goals of the proposed redesign of the Diary and Quarterly Interview surveys, as discussed in Chapter 1, are as follows: (1) reducing measurement error, (2) reducing burden and (3) increasing flexibility. The data requirements have to do with addressing the needs of the CPI (Casey, 2010) and the larger CE user community (Henderson et al., 2011).

Goals of the Redesign. The recommended procedures for the Diary and the Quarterly Interview survey address the goals of the redesign in multiple ways. For example, the proposed use of electronic and paper records could help address the interrelated goals of reducing measurement error and respondent burden. When the data are collected from records, respondents are relieved of the task of trying to remember their expenditures. Moreover, the records are likely to be more accurate than the respondents' recollections. The use of incentives may significantly boost the respondents' willingness to keep the appropriate records. The regular reminders by the interviewer should also improve respondent compliance.

We propose using multiple respondents in the Diary survey as a way to reduce respondent burden and reduce measurement error. Using multiple respondents, however, has a significant price tag. We recommend that BLS conduct a field experiment to evaluate the use of multiple respondents. However, even if the idea of multiple respondents is too expensive to implement, we still recommend that BLS develop procedures to elicit greater participation in the household. In the end, analysis will have to weigh the cost of this design against the decreases in measurement error. Total survey error models should be developed to assist in understanding the tradeoffs between reduction in error against reductions in the sample size that would be implemented to pay for the procedure.

We recommend reducing respondent burden in the Quarterly Interview survey by cutting the number of waves a CU is in sample from four to two, as well as increasing the time between waves. While this will increase data collection costs, it also significantly increases the precision on the annual estimates. As noted in the Quarterly Interview precision section, the reductions are

substantial. Further refinement of both the cost and simulation model discussed above is needed. The cost models were based on aggregate data for the current design, which hampered our ability to refine costs. In addition, further information is needed how the costs and reductions in sample size relate to one another.

For the Quarterly Interview survey, we propose collecting administrative data from utility companies, apartment lessors, and other sources. These data would substitute for respondent reports, reducing measurement error and respondent burden. Of course, collecting administrative data also has a significant cost associated with it. Unlike the elimination of two-panel waves, there is no clear cost offset to this component. An important evaluation question is how effective this component to the survey is to reduce measurement error with respect to total survey error. By investing resources in the administrative record followup, is the survey reducing measurement error more than it is increasing variability due to sampling? It is difficult to know without more research.

When viewed as a single survey program, the redesigned procedures address flexibility in several ways. The redesign allows respondents to use a number of different data collection methods that can be tailored to the respondent's daily routines. The challenge will be to develop a procedure that takes advantage of this flexibility without making the task overly complicated or to give respondents a way to avoid completing the task. The electronic record component is largely a passive method which only requires respondents to download files that have some basic information about their transactions. If a respondent agrees only to this, there is potential that significant gains can be made with respect to reducing measurement error and burden. As noted in Chapter 3, the electronic records cover a very large proportion of all expenditures—potentially over 70 percent. In our proposed redesign, the respondents are asked to download financial data files. They are also asked to save paper receipts. If they do not adequately download the files or keep the receipts, the interviewer conducts a recall interview to capture the needed data.

Our proposed redesign also offers the respondents flexibility. At some point, there may not be a need for two surveys. The two may merge into a single collection that primarily relies on relatively passive methods that can address the data requirements of the CE program. A merged survey was considered when developing the redesign options. This was rejected because it was not deemed feasible for the 4 to 5 year time frame that the redesign is targeting. But it would not be difficult to merge the two collections into a single survey if that was warranted.

Addressing CE Data Requirements. The CE must generate information in a way that is usable for the CPI. The expenditure data and demographics must be at the appropriate level of detail,

geographical distribution, and periodicity (Casey, 2010). These data have to be produced in a timely way and with sufficient sample size to generate precise estimates. The suggested redesign procedures were structured to maintain the current expenditure categories generated around the same timeliness as the current CE. We have assumed the same sample design as in the current survey. Consequently the level of geographic detail needed keeps the status quo. Similarly, the discussion above has provided information on how the designs might affect precision of the estimates. The redesigned surveys are more expensive, which affects costs and, in turn, could affect sample sizes. For purposes of meeting the requirements of the CPI there would need to be adjustments to either the budgets or cutting costs to meet these.

We have also kept constant the level of expenditure detail provided for the CPI. It was difficult for the redesign team to consider how to reduce this detail in a logical and meaningful way. If the use of records is increased dramatically, reducing detail may not be as critical. However, some consideration should be given to cutting this detail to both reduce error and burden.

The requirements of other CE users are discussed by Henderson et al (2011). These requirements range from the annual estimates to household expenditure patterns. These requirements are shown in Table 18, along with the impact of the recommended redesign. One requirement is that a complete picture of spending be provided for a particular CU. This is met in both the Diary and Quarterly Interview components by covering all expenditures. The Interview survey will not request respondents keep their receipts for food and alcoholic beverages, in an effort to keep this task manageable. But the recall interview will cover these expenditures in summary form, as the current survey does now. We have discussed the idea of creating rotating modules. These modules would collect data at a summary level for a certain class of expenditures, With the exception of food and alcoholic beverage expenditures, which come from the Diary survey, these modules would lead to some reduction in the precision of the estimates. Whether or not a module is used, a complete picture of spending is possible with the redesigned method.

This design does not provide this complete picture for an entire annual period. By restricting the panel to two waves, spaced 12 months apart, it would not be possible to see how annual expenditures occur for individual households. One could look at how expenditures change within any particular 12-month period by comparing households that are interviewed during different parts of the year.

Table 18. Data requirements for CE analysts and redesigned CE program

Requirement	Addressed by redesign?
Complete picture of spending	Yes. If modules are used on Interview, imputation might be required
Level of expenditure detail	Yes
Population	Yes
Sample size and geographic detail	May need to increase cost to meet current requirements
Time period and periodicity	Yes
Panel component	Annual estimates of change. Not possible to estimate at a CU level: (1) expenditures for an entire year; (2) change for intervals of less than 1 year
Integrated microdata	Yes
Income	Yes, But not directly addressed in recommendations
Socioecomic information	Yes
Assets and liabilities	Yes, but not directly addressed in recommendations

The one requirement that is not met at the same level as the current survey is the panel component. The current survey provides estimates of change on a quarterly basis, while the redesigned method does so on an annual basis. The design also reduces the precision of estimating quarterly change in the aggregate, since none of the CU's will overlap within any particular quarter. As noted above, this change also seems to lead to more costs, although it isn't clear how much, if any, given the reductions in variance under the redesigned survey. If the advantages of going to a 12-month panel are not viewed as important as measuring quarterly change, the design could use a quarterly panel, perhaps only using two, rather than four waves. This should reduce costs relative to the 12-month panel, but it will increase variances of annual estimates as well relative to the recommended design. Another alternative would be to adopt a hybrid approach that ran two different types of panels, one with a 12 month and another with a quarterly design. Of course this complicates the design even more than is currently the case with two surveys.

Priorities of the Proposed Redesign. While the proposed design has many advantages, the cost of the full design, as shown in Table 19, is significantly higher than the current design. Table 19 provides a summary of the additional costs, relative to the current design, associated with the primary redesign components.[4] We recommend the highest priority be given to the use of personal records in the collection of the information. This has the greatest long-term potential to address both the measurement error and burden placed on respondents. It is also a cost-neutral feature of the new design.

[4] For purposes of discussion, we do not separate the cost of the incentives from other costs associated with the proposed interview design (e.g., using two, rather than four, waves of data collection). We believe this is a conservative assumption given the evidence that the incentives proposed are cost neutral and may even save the survey money.

Table 19. Additional costs of proposed design relative to current design (in millions)

Method to reduce burden and measurement error	Diary	Interview	Total
Use of personal records	n	N	n
Administrative Records	.8	10.2	11.0
Reduction in CE Interviews	na	7.6*	7.6
Multiple Diary Keepers	2.4		2.4
Enhanced Recall Interview	n	N	n
Total	3.2	17.8	21.0

n= not a significant cost factor

There are two major components of additional costs associated with the proposed design. One is the proposed interview structure of two, rather than four, panel waves. This accounts for approximately $7 million of the $21 million in additional costs. This is a 40 percent increase in the cost of conducting interviews relative to the current Quarterly Interview survey. This does not account for the significant cost savings one could achieve by cutting sample size. For example, we have shown that cutting sample size on the proposed design by 20 percent would achieve similar levels of precision in the annual estimates as the current design if a 3-month reference period was used for all expenditure items. One could cut as much as 40 percent and still achieve similar precision for many of the items if some were shifted to a 12-month reference period. Research into whether this is possible should be part of the development process when developing and evaluating the recall interview.

The primary motivator to reduce the number of panel waves is to cut respondent burden. As noted above, this should be a high priority for the redesign. However, the proposed design has the analytic disadvantage of not providing annual estimates of expenditures for individual CUs for all expenditures. One could theoretically achieve a similar reduction in burden, while maintaining the current four-wave panel design in other ways. For example, one could significantly reduce the number of expenditures for which extensive detail is collected. Such a design would develop three or four broad classes of expenditures covered in the recall interview in either summary or detailed form. These classes might include expenditures on apparel, health care, and entertainment. The level of detail collected for each module would be rotated either across households or across panel waves. For example, some households may not be asked detailed questions about apparel, while other households would not be asked detailed questions about health care. To achieve similar reductions in burden as the proposed design, approximately half, or slightly more than half, of the current Quarterly Interview survey would be administered in detailed form at any particular point in time.[5]

[5] The current Quarterly Interview design conducts four 3-month interviews and one bounding interview for an individual CU. The proposed design conducts two 3-month recall interviews and two bounding interviews for an individual CU.

The use of modules has disadvantages associated with it. One is that it complicates analysis of total household expenditures. A second issue is that it will reduce the precision of the annual estimates. If a module collects the detail needed for annual estimates on 60 percent of the interviews, the precision of the annual estimate would be reduced by approximately 40 percent, all other things being equal. Because of time and resource constraints, we did not fully examine the possibilities of this type of design. But it should be considered as part of the redesign efforts.

A second major component of additional costs of the proposed design is the administrative record survey. The survey proposed for the Quarterly Interview survey accounts for $10 million of the additional $21 million. The cost represents a fairly extensive administrative survey which would collect data on many different types of retail purchases. The costs are based on the provider survey conducted for the Medical Expenditure Panel survey. In this case, respondents may go to multiple providers, depending on the type of health care expense. It would be possible to significantly reduce the cost of the proposed administrative survey by concentrating on utilities, such as electricity, gas and water. There will be many fewer companies that would need to be contacted when compared to contacting retailers. Based on our recent experience with the Residential Energy Consumption Survey (RECS), this could reduce the cost of $10 million to as low as $3 million. This cost might be further reduced by restricting the collection of these data for a portion of the sample. If 12 months of data are collected at the same time, it would be possible to reduce the number of households for which the survey is conducted. For example, under the proposed design, shifting the utility items to a 12 month reference period reduced the variance by 56 percent. This implies that significant reductions in sample might be possible once shifting to this type of administrative source.

Variations on the Proposed Redesign. The variations around the proposed design should retain the personal record components. This potentially addresses both measurement error and burden without a significant increase in costs. It is also recommended that the use of multiple diary keepers be retrained, at least for serious consideration in a field experiment. The added expense of this procedure of $2.4 million is significant for the diary. However, the evidence on this change indicates it could lead to significant reductions in total survey error. It also reduces burden on the main diary keeper. Similarly we recommend all designs include an incentive for the Diary and the Quarterly Interview Surveys. The evidence suggests this will result in an increase in response rates and a reduction in field costs.

One lower cost variation on the proposed design would be to narrow the scope of the administrative survey to the collection of utility information. As noted above, this cuts the cost of the administrative survey by approximately 70 percent. This variation would maintain the

recommendation of cutting the number of panel waves from four to two. By lengthening the reference periods for key items, it might be possible to cut the sample of the survey, as discussed in the section on precision for the Quarterly Interview survey. If it is assumed that sample could be cut by 20 percent, then the increase in the cost of interviewing on the Quarterly Interview Survey would be approximately $3.8 million. This assumes a proportionate reduction in cost relative to reductions in sample size. Combined with the reductions in the scope of the administrative survey, this variation on the proposed design increases costs by approximately $9.4 million[6] compared to the $21 million.

A second variation would cut costs further by maintaining the current four-wave panel for the Quarterly Interview Survey. As with all of the recommended designs, this second variation would include multiple diary keepers ($2.4 million). It would also include the collection of utility data using an administrative survey ($3 million). The advantage of this design is that it provides annual expenditure information for individual CUs, which is not something the proposed design can do. The disadvantage of this approach is that the only measures that directly reduce burden are the use of personal records and collection of administrative data on utilities. CUs are retained in panel and are asked to complete 4 waves of recall interviews.

[6]$3 million for the administrative record survey, $2.4 million for multiple diary keepers, and $3.8 million for the revised interviewing structure.

Development and Evaluation of the Proposed Design 6

This section presents a series of research and development (R&D) activities to identify the most promising approaches to redesigning the CE program. The section first presents R&D activities that apply to both the Diary survey and the Quarterly Interview survey. It then presents R&D activities that pertain to each component.

Common Elements

Some of our proposed data collection procedures are common to the Diary survey and the Quarterly Interview survey. For example, we propose that both surveys obtain expenditure data from retailers with loyalty card programs. We propose that in both surveys, respondents be asked to save and scan receipts, or to mail in their receipts.

Software Development. Our recommended redesign would require that new software be developed. R&D activities would then address these issues:

- Financial software like Mint can download data files for checking, credit card, and other accounts from the web sites of banks and financial companies. Users choose financial institutions from a list, and enter their account numbers and passwords. The software then automatically downloads the files. Can similar software be developed for the CE program, perhaps using the Yodlee SDK? Could the software then automatically transfer the files electronically to the central repository? Will CE respondents be willing to use this software, or will many find it burdensome, or perceive a risk to their privacy? Will some respondents prefer to download and email these files themselves, without using any special software?

- In our proposed CE redesign, respondents would be asked to save their receipts and to complete a short form describing purchases that had no receipts. They would then scan their receipts. The scan files would be sent electronically to a central repository. OCR software would then turn the scans into text. The system would then extract the individual purchases from the text. All of this would occur automatically; the respondents would need only to start the software and scan the receipts and forms. Existing software such as Neat Receipts has some of this functionality, although it cannot extract the individual items purchased. Can all this software be developed so that CE respondents can use it without problem? Will the software successfully extract data about individual purchases?

- The system would match the purchases identified from the receipts with the purchases identified from the downloaded data files for the respondents' credit card, debit card, checking, PayPal, and other accounts. Existing software products like Mint and Moneydance can accomplish this matching. The matching is needed so that the web survey would ask about these expenditures only once. Can software be developed with this functionality for the CE program?

- Once the system identified the individual expenditures, it would generate a web survey to ask the respondents the details about each one. Can software be developed to generate a usable survey? Will respondents be willing to complete the survey?

- Potentially, smartphone apps could be created that enable respondents to enter notes about purchases, or to photograph purchases or bar codes. The web survey would prompt the respondents to enter more data about the purchases. Could such a system be developed, including the smartphone apps? Would many respondents use it successfully?

The R&D activities for the software needed in our proposed redesign would have several steps:

- Prototyping;
- Testing for function and usability;
- Revisions and pilot testing;
- Further testing and revisions; and
- Field testing.

Development of Procedures and Materials. Our proposed redesign would require a number of hard-copy materials, including the consent forms, the short forms that respondents complete to record purchases that had no receipt, and envelopes and boxes for respondents to store receipts, and printed instructions. The development of these hardcopy materials should include cognitive interviews and usability tests, to address issues such as the following:

- Are the printed instructions easy for respondents to read, understand and use?
- Do the respondents tend to have any specific difficulties following the instructions?

- Are members of the household other than the respondent generally willing to read the instructions, and take the time to fully understand them?

- Can respondents easily use the hardcopy forms to record purchases which have no receipt? What is the most usable format for these forms? Should they be designed so that respondents record one purchase per form, or should one form cover a single day, or is some other formatting best?

The initial testing for these hardcopy materials could be a set of interviews in which respondents are asked to use and react to prototypes. In later rounds of testing, respondents could use the materials at home, perhaps in concert with other household members. The testing would assess the respondents' ability to understand and use the materials.

Small Field Pilots. Once initial versions of the hardcopy materials and the software were developed, small pilot tests could be run to address issues such as these:

- Can respondents follow instructions on the use of the different methods to collect expense information (such as downloading files, and scanning receipts)?

- If CUs were to have multiple respondents, can the principal CU respondent successfully explain the CE survey procedures to other CU respondents?

- Do other CU respondents follow the instructions? What issues arise when they are asked to follow these instructions?

- Do respondents prefer to be given a choice in data collection methods, or do they prefer to be prescribed a method? Why?

- What types of technical problems do respondents have when using the software?

Respondents would be recruited for these field tests with the understanding that the tests would involve two contacts. During the first contact interviewers would provide verbal and printed instructions. A second in-person visit would be conducted two weeks later to end the data collection period. An interviewer would call every 3 or 4 days over the 2-week period to assess the respondents' ability to download financial data files, save receipts, and follow the other procedures. During the second visit, the interviewer would debrief the respondent and other members of the household to learn the problems they encountered and their views of the procedures.

Procedures to Assign Methods of Reporting. The procedures afford respondents a number of choices for reporting their data. They could download financial data files, scan receipts, complete and scan paper forms, or mail in paper receipts and forms. Respondents could possibly also provide

some expenditure data using smartphones, by completing on-screen forms or recording voice memorandums about individual purchases, or photographing their purchases.

By giving the respondents a range of options, our proposed redesign might tailor the survey procedures to the needs and preferences of the respondents. However, we cannot argue that maximum flexibility will necessarily result in a higher response rate or better quality data. Giving respondents choices for how they can respond can actually result in a diminished response rate (Messner et al., 2009; Dillman et al., 2007). CE respondents should be encouraged to provide data using electronic methods, which may be more efficient and less burdensome than methods that rely on paper. Recent research on mode choice (Christian et al., 2011) suggests that respondents should be offered choices in a sequential manner. If that is true, then respondents should be asked to download financial files and scan their receipts and forms, and not offered any alternatives, such as mailing in paper receipts, unless they objected to the electronic methods.

On the other hand, the CE program is unique in many ways; prior research on mode choice may not completely apply to the CE program. Perhaps respondents should be presented with the electronic and paper alternatives and allowed to choose. Perhaps interviewers could ask respondents about their use of technology, and their habits regarding saving receipts, and their use of smartphones. The data collection method could then be tailored to assign a methodology that would maximize the use of electronic and paper records.. We recommend a field test to determine the best way to build flexibility into the data collection methods.

Collection of Administrative Data. Several issues are unresolved about the use of administrative data in the CE program. Some of the issues that must be investigated to determine the role of administrative data include the following:

- Will retail, utility and mortgage companies agree to provide a respondent's information if they receive a signed consent?
- What are the major barriers to obtaining the information from these companies?
- What is the most efficient method for gaining the companies' cooperation? Is it contacting the companies' central business offices?
- Are there ways that BLS can help motivate the companies to provide data, such as through publicity, or a payment?

- Can information obtained from retail, utility, and mortgage companies be easily compared with a respondent's expenditure reports?
- How much personnel cost is involved with obtaining information from these companies?

A field test could investigate whether many respondents are willing to sign consent forms to allow the interviewers to obtain these data, the conditions under which retail, utility, and mortgage companies would provide data, and the extent to which these data can reduce measurement error. These data potentially could serve as a substitute for respondent reports. The data potentially could be used for adjustments to the respondents' data, or to fill in data about expenditures that are missing from the respondents' reports.

Comparison to Current Production Survey. Once the designs of the surveys have been finalized, there should be a test that compares the redesigned methods for the CE surveys with the methods of the current production CE surveys. The field experiments described below can help compare the redesigned methods with the current methods. However, these studies would not use a nationally representative sample. Nor would they use a full data collection cycle. A comparison using representative samples and a full data collection cycle would provide a more rigorous test of the new, redesigned data collection methods.

Several other redesigns of large federal surveys have involved comparisons between the new and old methodologies. For example, the redesign of the Current Population Survey (CPS) included a test to compare the redesigned methods with the current methods. This test provided data to help refine the new survey methods. However, the new procedures evolved over the life of the data collection period, making comparisons with the current methods difficult to interpret.

The test of the redesigned National Crime Victimization Survey (NCVS) in the early 1990s also included parallel surveys. The results allowed statisticians to calculate estimates while the survey data collection methods evolved to the new, redesigned ones (Kindermann, Lynch, & Cantor, 1997).

The field experiment for the CE Interview survey should involve a national sample, or at least a sample that could provide results that could be generalized to a specific geographical area, so that the results using the redesigned data collection methods could be compared with the results using the current methods, and so that the results could be benchmarked to estimates from the American Community Survey, the PCE, or other national surveys. In the field experiment, the redesigned

survey should be administered over at least one full collection cycle, which would be one calendar year.

Development and Testing of the Diary Survey

R&D activities for the Diary survey would have several components: (1) development of the recall interview, (2) pilot test to evaluate the data collection procedures, and (3) a field experiment.

Developing the Recall Interview. In our proposed redesign of the CE Diary survey, respondents will be asked to complete an in-person recall interview at the end of the 2-week data collection period. This interview would employ visual aids such as Event History Calendars, and receipts.

R&D activities should include assessing the value of various cuing strategies for helping respondents remember expenditures during the recall interview. These studies could start with a card sorting task that asked respondents to categorize different types of expenditures into different piles. (Westat will soon conduct such a card sorting research study under contract with BLS.) The card sorting task will reveal how respondents themselves group their expenditures. The results of this research will suggest how the recall interview can best be structured.

A second study could evaluate an approach to data collection using an Event History Calendar, a method that has been used successfully in other surveys (e.g., Belli, Shay, & Stafford, 2001). An interviewer organizes an Event History Calendar around events during the prior week that were significant for the respondent. The interviewer then uses these significant events to help cue the respondent to recall expenditures.

Respondents in the evaluation study might save their receipts for a 1-week period. The respondents, without referring to these receipts, would be asked to report their expenditures. The receipts could be used to assess the accuracy of the data that the respondents provide in the recall interview.

Pilot Study. A pilot study could address the following questions about the Diary survey:

- Is it feasible to obtain records from retail, utility, and mortgage companies to compare against respondent reports? Will respondents be willing to sign consents and will these companies be willing to provide the data?
- What proportion of respondents is willing to download financial data files, scan receipts and forms, and email these data? What proportion is willing to mail in hardcopy receipts and forms?
- Are any respondents unwilling to provide data either electronically or on paper, but still willing to participate in a recall interview after 1 week? After 2 weeks?
- Are respondents generally willing to provide consent to allow the interviewers to access loyalty card information?
- How should interviewers tailor the data collection method to the respondent? Should the interviewer allow the respondent to choose a method, or should the interviewer prescribe a method?
- How effective is the recall interview in obtaining accurate and complete data?
- What types of problems do respondent have with using the software, scanner, smartphone apps, and software for downloading financial files?
- What is the participation rate from CU members who are not the principal respondent, when CUs have multiple respondents?
- What is the reaction of respondents to the various incentives?
- What is the perceived burden of the procedures?
- Are interviewers able to effectively monitor the expenditure data that the respondents mail or email to the central repository?
- Is the system able to effectively reconcile expenditures from the receipts and the financial data files?

In the pilot study, approximately 200 households would be assigned to each of four groups which differ by the number of people interviewed in the household (one respondent vs. all adults, with a principal respondent) and the approach taken to determine the response method (interviewer assignment vs. respondent choice). This design will provide an initial assessment of how the data collection procedures affect data quality and cost.

The respondents' expenditure reports would be verified by comparing them to retailers' administrative records. The feasibility of this method, of course, is dependent upon the

establishment of a relationship between BLS and retailers. Ideally, several different types of stores such as grocery chains and "big box stores" can be recruited to provide purchase histories for specific loyalty card numbers. Respondents might be recruited from customers listed in the loyalty card database.

There are alternative, less direct methods to sample store customers. For example, BLS could actively recruit customers through advertisements in newspapers or web sites like CraigsList asking for persons who have loyalty cards from particular retail chains. Respondents could be recruited using intercept methods, such as when they exited the store.

Field Experiment. The next step will be to conduct a field experiment that addressed all the questions from the pilot study, plus these:

- What is the most effective combination of incentives to maximize the response rate and the respondents' use of electronic data collection methods (downloading financial data files, saving and scanning receipts and forms)?
- Which is most effective: asking a single respondent to provide data or asking every adult in the CU?
- What is the level of burden associated with the redesigned procedure?
- What is the response rate and associated nonresponse bias of the redesigned procedures?
- Can respondents effectively use the different software and hardware applications? What are the problems they encounter?
- How do the new procedures compare to the current design, with regard to data quality and costs?

The design is illustrated in Table 20. The test would include a total of nine experimental conditions: two response methods (respondent choice vs. methods based on respondent habits) would be crossed with two conditions for respondent selection (single respondent vs. multiple respondents with one principal respondent). Two incentive conditions would be nested within each respondent selection method. The current data collection method would be included as an additional condition.

Table 20. Experimental design for field test for the Diary survey

Response method	New design – respondent selection				Current design
	Single respondent		Multiple respondent		
	$50	$70	$50/$10	$80/$20	9
Method 1	1	2	3	4	
Method 2	5	6	7	8	

The respondents included in the test would be ones whose reports could be compared with data from retailers. By including the current collection methods as a treatment, the field experiment would allow for comparisons with the experimental procedures.

Sample. The sample would be drawn from selected geographical areas that spanned urban and rural populations. As with the pilot test described above, the sample should ideally be drawn from retailers' loyalty card databases, so that the participants' reported expenditure data can be compared with the database for validity.

Evaluation Measures. Multiple outcome measures are needed to address the research questions. The outcome measures would include the following:

- Mean level of expenditures;
- Proportion of reported expenditures that match retailer information in the loyalty card database;
- Proportion of expenditures that the respondents provide using records, rather than memory;
- Cost of data collection per CU;
- Non-response and bias attributable to non-response;
- Respondents' perceived level of burden;
- Time required to complete interviews;
- Proportion of respondents requiring the recall interview because they failed to provide adequate data via mail or email;
- Proportion of expenditures reported during recall interview; and
- Level of respondent satisfaction with the procedures overall, and specifically with the software and scanning hardware.

The direct measure of quality will be a match between the survey and retailer records of selected purchases. Other, less direct, measures of quality would be the mean expenditure levels, the percentage of respondents using records and the percentage of expenditures that are based on records. A debriefing interview will ask respondents about the process such as satisfaction with different aspects of the process and any problems they had with the process.

Sample Sizes and Statistical Power. There are three experimental factors proposed for the above field test, in addition to testing the current design. Table 21 provides the power for different sample sizes to test hypotheses related to differences in the response rate. The sizes shown represent the number in each of the comparison groups. For example, for a sample size of 800 in each group, there would be 80 percent power to detect a 7 percentage point different when one of the percentages was around 50 percent. If one wanted to be able to conduct this test for comparable cells in the experimental design, this would put the total sample size to around 7,200 (9 cells x 800). One could also design the study assuming no interactions between some of the experimental factors, which would reduce the sample size.

A second line of research would be to examine the match rates between expenditures in the administrative records and the survey reports. For example, a question such as "What percent of respondents reported the expenditures listed in the administrative record?" The power listed in Table 22 assumes a design effect of 2, considering the geographic clustering of the sample. The rows correspond to a 60 percent response rate for the sample sizes listed in Table 21. For example, a sample of 800 corresponds to approximately 500 completes (800 x .6). With 500 completes and a percentage of 70 percent, there would be 70 percent power to detect a difference of 10 percentage points. With 1,000 per group the power goes up considerably.

The third set of questions to be asked would compare the mean expenditures across the different experimental treatments. Since it is suspected that many different types of expenditures are under-reported, one point of interest would test if the mean of the experimental treatments were higher. Power calculations were made based on the published standard errors for the Diary survey for selected expenditures. Table 23 provides the power for different sample sizes. The effect sizes represent a 10 percent or 20 percent increase in the mean, using the published estimate for that category. With the large variation in expenditures, considerable sample size is needed to precisely test for differences. For example, 1,000 completed interviews is needed to have respectable power (80%) for a most of the expenditures listed, except alcoholic beverages and eggs, both of which have high standard deviations.

Table 21. Power for difference in response rates by sample size, level of response rate and size of difference

	p=50%		p=70%		p=80%	
Sample size	2+	7	2	7	2	5
800	12%	80%	14%	87%	16%	72%
1600	20%	98%	23%	99%	29%	95%
3300	37%	100%	43%	100%	53%	100%

+ Assumed Percentage Point Difference in Response Rates.

Table 22. Power for difference in match rates by number of completes, level of match rate and size of difference*

	p=50%		p=70%		p=80%	
Completes	4+	10	4	10	4	8
500	14%	61%	16%	70%	19%	62%
1,000	24%	89%	28%	94%	35%	91%
2,000	43%	99%	50%	100%	61%	100%

+Assumed percentage point difference in match rate.

*Assumes a design effect of 2.

Table 23. Power for difference of means for selected expenditures, number of consumer units completing a diary and effect size*

	n=500		n=1,000		n=2,000	
Expenditure type	10%+	20%	10%	20%	10%	20%
Food at home	24%	72%	43%	95%	72%	100%
Cereals and Cereal Products	17%	51%	29%	80%	51%	98%
Eggs	13%	39%	22%	67%	39%	92%
Other Food at Home	19%	58%	34%	87%	58%	99%
Food Away From Home	22%	65%	38%	91%	65%	100%
Alcoholic Beverages	7%	15%	10%	26%	15%	46%

*Power calculated for a two sided significance test at p<.95.

+ Effect size = percent difference between the two estimates of expenditures.

It should be noted that these estimates of power assume no change in the design relative to the national survey. In a field test it would be possible to gain efficiencies through stratification and blocking the sample (e.g., assign different treatments to households on the same block). It also may not be necessary to extensively weight the data. This should lead to a more precise test of the hypotheses of interest. Similarly, one could adjust the significance levels from 95 percent to 90 percent, which also increases power. Nonetheless, these calculations indicate that significant resources will be needed to test design elements for the CE. It is important to note that when the

CE was first being designed, Neter and Waksberg (1964) conducted an experiment of approximately 5700 households, each interviewed 4 times over 18-month field period.

Development and Testing of the Quarterly Interview Survey

The Quarterly Interview survey will involve an initial in-person contact with the household in which the interviewer would provide instructions, obtain consents and conduct a bounding interview. The interviewer would then monitor the extent to which the respondent provided data by mail or email. The interviewer would call during that time, and after 1 and 2 months, to encourage the respondents to continue to send in expenditure data. The interviewer would conduct the second in-person visit after 3 months. This interview would involve conducting a recall interview. A second round of contacts would occur 12 months after the initial contact. The purpose of these contacts would be to re-orient the CU with the data collection task and to conduct a bounding interview by telephone. The interviewer would make interim prompting calls over the first two weeks, and after 1 and 2 months. The final contact at 3 months would be in person to conclude the data collection and conduct a recall interview.

The Recall Interview. The recall interview for the Interview Survey can be developed based on the results of the field test of the Diary survey. In particular, the results of the card sorting task study could guide the structure of the interview. These results will suggest how respondents tend to categorize expenditures, and can point to the optimum cuing strategies. An EHC procedure should be developed as well.

The optimum reference periods for the Quarterly Interview survey must be determined. Research needs to settle the issue of which expenditures are best handled with a 1, 3, or 12-month reference period.

That research could use a design in which different groups of CUs were asked to provide data about various expenditures using various reference periods. For example, one group of CUs could be asked to report on expenditures in the domain "personal services" for a 1-month period. That same group could then be asked to report for 3 months and then for 12 months. Another group of respondents could be asked to report on "personal services" expenditures in the reverse order: first 12 months, then 3 months and then 1 month. The results of this research would suggest he types of

errors respondents tend to commit for various types of expenditures using the three reference periods.

Pilot Study. We recommend a pilot test of the proposed redesigned procedures. The interviewers would have four contacts with the CUs. In the first two, the interviewers would collect expenditure data for a 3-month period. In the second two, the interviewers would collect expenditure data for a second 3-month period that started 12 months after the first contact.

The questions that this pilot study would address are the same questions that the pilot study for the Diary survey addressed. Moreover, this pilot study would address two questions unique to the Interview survey:

- How effective is the recall interview for various expenditure categories when different reference periods are used?
- Is there a difference in response rate and data quality between the first and second year of the Quarterly Interview survey?

Again, the sample of CUs should ideally be recruited from retailers' loyalty card database records so that the respondents' expenditure reports can be compared with data from the database.

Samples of approximately 200 CUs for each of eight groups would be recruited. These groups would vary by: (1) number of people interviewed in the household (one respondent versus all adults), (2) the approach taken to determine the response method (the interviewer assigns the response method versus the respondent chooses), and (3) the reference period assigned to specific expenditure categories (all 3-months versus 1, 3, and 12 months). Decisions on the allocation of different expenditure/reference period combinations would be made based on the results of the development work described above.

The methods of the pilot study for the Quarterly Interview survey would be identical to those described above for the pilot study for the Diary survey.

Field Experiment. A field experiment would reveal how redesigned procedures compare to the current ones. The complexity of the design of the field experiment would depend on whether or not the results of the pilot test strongly suggested that specific data collection procedures were desirable, and did not require further evaluation. One design would be identical to that tested for the Diary

survey (see Table 19). If it is desirable to experiment with different reference periods, the selection method could be dropped from the design and replaced with a reference period factor (Table 24).

Table 24. Design for field test for interview survey that includes reference period as experimental factor

Reference Period Group	New design – respondent selection				Current design
	Single respondent		Multiple respondent		
	$50	$70	$50/$10	$80/$20	9
Group 1	1	2	3	4	
Group 2	5	6	7	8	

Sample Design and Evaluation Measures. The sample design and evaluation measures would be identical to those of the Diary field experiment, described above.

Sample Size and Power Calculations. There are three experimental factors proposed for the above field test, in addition to testing the current design. Many of the analytical questions to be addressed for the Quarterly Interview survey are similar to those discussed for the Diary. Tables 20 and 21 provide the power for examining hypotheses related to differences in response rates and the rate of matches between administrative records and survey reports.

The third set of questions to be asked will compare the mean expenditures across the different experimental treatments. To compute the power for these estimates, variances and standard deviations were calculated using the 2009 data available on the public use data-set. A single collection quarter was used to simulate what would be collected during the field test. The means for the interview survey have higher variability (per CU) than the Diary. Therefore the power is not as good for comparable sample sizes. Table 25 provides power for the test of whether the means differ by 10 percent, 20 percent or 30 percent. Except for electricity expenditures, none of the tests achieve the 80 percent level when testing for effect sizes of 10 percent or 20 percent. Even with 2,000 completed interviews, not all of the comparisons are at this standard for 30 percent effect sizes.

Table 25. Power for difference of means for selected expenditures, completed interviews and effect size*

Expenditure Type	N = 500			N=1,000			N=2,000		
	10%+	20%	30%	10%	20%	30%	10%	20%	30%
Electricity	46%	96%	100%	74%	100%	100%	96%	100%	100%
Apparel and services, Men, 16 and over	7%	13%	22%	9%	20%	40%	13%	36%	67%
Apparel and services, Women, 16 and over	9%	21%	42%	13%	38%	70%	21%	65%	94%
Vehicle Maintenance and repairs	9%	20%	39%	12%	35%	66%	20%	61%	92%
Medical supplies	5%	7%	9%	6%	9%	13%	7%	12%	22%

*Power calculated for a two sided significance test at p<.95.

+ Effect size = percent difference between the two estimates of expenditures.

References

Battaglia, M.P., Shapiro, G., and Zell, E.T. (1996). *Substantial bias may remain when records are used in a telephone survey.* Presented to the American Statistical Association. Available at: http://www.cdc.gov/nchs/data/nis/data_collection/battaglia1996.pdf.

Belli, R.F., Shay, W.L., and Stafford, F.P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly,* 65, 45-74.

Belli, R.F., Stafford, F., and Alwin, D. (2008). *Calendar and time diary methods in life course research.* Los Angeles, CA: Sage.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., and Yamamoto, K. (1992). An experiment in monetary incentives. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/Proceedings/.

Biderman, A.D., Cantor, D., Lynch, J.P., and Martin, E. (1986). *Final report of research and development for the redesign of the National Crime Survey* (Prepared for the Bureau of Justice Statistics). Washington, DC: Bureau of Social Science Research, Inc.

Blank, R.M., and Barr, M.S. (Eds.). (2009). *Insufficient funds: Savings, assets, credit, and banking among low-income households.* New York, NY: Russell Sage Foundation.

Bogen, K., Moore, J.C., and Marquis, K.H. (1994). Can we get respondents to use their personal income records? *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/proceedings/papers/1994_220.pdf.

Bradburn, N. (2010). *Recall period in Consumer Expenditure Surveys Program.* Paper presented at the Consumer Expenditure Survey Methods Workshop. Available at: http://www.bls.gov/cex/methwrkshp_pap_bradburn.pdf.

Brick, J.M., Kalton, G., Nixon, M., Givens, J., and Ezzati-Rice, T. (1999). *Statistical issues in a record check study of childhood immunization.* Presented to the Federal Committee on Statistical Methodology. Available at: http://www.fcsm.gov/99papers/murphy.pdf.

Brick, J.M., and Williams, D. (2009). *Reasons for increasing non-response in US households.* Presented at the Workshop of the Committee on National Statistics, Washington, DC.

Cantor, D. (2010). *Discussion of plans for designing the recall period for the Consumer Expenditure Interview Survey.* Presented at the Consumer Expenditure Survey Methods Workshop. Available at: http://www.bls.gov/cex/methwrkshp_pap_cantor.pdf.

Casey, W. (2010). *CPI requirements of CE.* Available at: http://www.bls.gov/cex/ovrvwcpirequirement.pdf.

Church, A.H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57, 62-79.

Clifford, S. (2011, August 7). Shopper receipts join paperless age. *New York Times*, p. B1.

Creighton, K.P., King, K.E., and Martin, E.A. (2007). *The use of monetary incentives in Census Bureau longitudinal surveys* (Research Report Series, Survey Methodology). Washington, DC: U.S. Census Bureau.

Dijkstra, W., Smit, J.H., and Ongena, Y.P. (2008). An evaluation study of the event history calendar. In R.F. Belli, F. Stafford, and D. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 257-275). Los Angeles, CA: Sage.

Dubreuil, G., Tremblay, J., Lynch, J., and Lemire, M. (2011). *Redesign of the Canadian Survey of Household Expenditures.* Presented at the Household Survey Producers Workshop, Bureau of Labor Statistics. Available at: http://www.bls.gov/cex/hhsrvywrkshp_dubreuil.pdf.

Duly, A., Garner, T., Keil, E., Reyes-Morales, S., and Wirth, C. (2003). *The Consumer Expenditure Survey and AC Nielsen Survey: A data comparison study* (unpublished internal document). Washington, DC: Bureau of Labor Statistics, U.S. Department of Labor. Cited by Garner, McClelland, & Passero (2009).

Edgar, J., Davis, J., Spell, S., Verlander, R., and Wetzel, G. (2006). *Individual diary feasibility report* (Census Bureau report). Washington, DC: U.S. Census Bureau.

Einav, L., Leibtag, E., and Nevo, A. (2008). *On the accuracy of Nielsen Homescan data.* Washington, DC: Economic Research Service, U.S. Department of Agriculture.

Eltinge, J. (2011). *Potential topics for study in evaluating uses of external administrative data for the Consumer Expenditure Surveys.* Discussion presented at the first Panel on Redesigning the BLS Consumer Expenditure Survey. Available at: http://www.bls.gov/cex/redpanl1_eltinge.pdf.

Eppig, F.J., and Chulis, G.S. (1997). Matching Medicare Current Beneficiary Survey and Medicare data: The best of both worlds. *Health Care Financing Review*, 18, 211-229.

Ezzati-Rice, T.M., Zell, E.R., Massey, J.T., and Nixon, M.G. (1996). Improving the assessment of vaccination coverage rates with the use of both household and medical provider data. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/proceedings/papers/1996_054.pdf.

Filippucci, C., and Ferrante, M.R. (1997). Individual diaries and expense documents in the Italian Consumer Expenditure Survey. *Journal of Official Statistics*, 13(2), 113-121.

Foster, K., Meijer, E., Schuh, S., and Zabek, M.A. (2011). *The 2009 Survey of Consumer Payment Choice.* Boston, MA: Federal Reserve Bank of Boston. Available at: http://www.bos.frb.org/economic/ppdp/2011/ppdp1101.pdf.

Fox, E.J., and Sethuraman, R. (2006). *Retailing in the 21st century.* New York, NY: Springer.

Garner, T.I., Janini, G., Passero, L., and Vendemia, M. (2006). The CE and the PCE: A comparison. *Monthly Labor Review*, pp. 20-46. Available at: http://www.bls.gov/opub/mlr/2006/09/art3full.pdf.

Garner, T.I., McClelland, R., and Passero, W. (2009). *Strengths and weaknesses of the Consumer Expenditure Survey from a BLS perspective.* Presented at the Summer Institute, National Bureau of Economic Research, Cambridge, MA.

Grootaert, C. (1986). The use of multiple diaries in a household expenditure survey in Hong Kong. *Journal of the American Statistical Association*, 81(396), 938-944.

Henderson, S., Passero, B., Rogers, J., Ryan, J., and Safir, A. (2011). *Consumer Expenditure Survey (CE) data requirements.* Available at: http://www.bls.gov/cex/cedatarequirements.pdf.

Hicks, W., Connor, S., Ward, P., et al. (2011). *An experiment evaluating different amounts of respondent monetary gifts.* Presented to the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ.

Holbrook, A., Green, M., and Krosnick, J. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67, 79-125.

Hubble, D., and Wilder, B.E. (1988). Preliminary results from the National Crime Survey CATI experiment. *Proceedings of the Survey Research Methods Section, American Statistical Association.*

Jackson, J. (2003, October 1). Card games: Should buyers beware of how supermarkets use "loyalty cards" to collect personal data? *Baltimore City Paper.* Available at: http://www.joabj.com/CityPaper/031001ShoppingCards.html.

Kashihara, D., and Wobus, D. (2007). The effect of reporting aid use on the accuracy of household-reported medical expenditure data. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/Proceedings/.

Kaufman, P. (2007). Strong competition in food retailing despite consolidation. *Amber Waves* (Economic Research Service, U.S. Department of Agriculture), 5(1), 4. Available at: http://www.ers.usda.gov/AmberWaves/February07/Findings/Strong.htm.

Kim, A. (2001). *Taking the poor into account: What banks can do to better serve low-income markets.* Washington, DC: Progressive Policy Institute. Available at: http://www.ppionline.org/documents/Banks_080601.pdf.

Kindermann, C., Lynch, J., and Cantor, D. (1997). *Bureau of Justice Statistics, National Crime Victimization Survey: Effects of the redesign on victimization estimates* (NCJ-164381). Washington, DC: Office of Justice Programs, U.S. Department of Justice. Available at: http://bjs.ojp.usdoj.gov/content/pub/pdf/ERVE.PDF.

Kojetin, B., and Miller, L. (1993). Estimating the accuracy of proxy responses at the dyadic level. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/Proceedings/.

Lessler, J. (1987). *Comments and ideas for the BLS research agenda for the Consumer Expenditure Survey and Current Population Survey* (Report to the Bureau of Labor Statistics). Available at: http://www.bls.gov/cex/cesrvymethslessler1.pdf.

Marquis, K.H. (1995). *The SIPP measurement quality experiment and beyond: Basic results and implementation.* Presented at the Census Bureau Annual Research Conference, Arlington, VA.

Martin, E., Abreu, D., and Winters, F. (2001). Money and motive: Effects of incentives on panel attrition in the Survey of Income and Program Participation. *Journal of Official Statistics,* 17, 267-284.

McKernan, S.M., and Sherraden, M. (Eds.). (2008). *Asset building and low-income families.* Washington, DC: Urban Institute.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research,* 20, 431-440.

Mint.com. (2011). *How to: Track cash spending and checks.* Available at: http://www.mint.com/blog/how-to/track-cash-spending-and-checks/.

Moneydance. (2011). *Balancing accounts using the Reconcile tool.* Available at: http://moneydance.com/userguide-contents/balancing%20accounts%20using%20the%20reconcile%20tool.html.

Moore, J.C., Marquis, K.H., and Bogen, K. (1996). *The SIPP Cognitive Research Evaluation Experiment: Basic results and documentation.* Available at: http://www.census.gov/sipp/workpapr/wp9601.pdf.

Neter, J., and Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 18-55.

The Nielsen Company. (2011). *Consumer panel and retail measurement.* Available at: http://www.nielsen.com/us/en/measurement/retail-measurement.html.

NPD Group. (2011). *Products and services: Retail tracking.* Available at: http://www.npd.com/corpServlet?nextpage=retail-tracking_s.html.

Perloff, J.M., and Denbaly, M. (2007). Data needs for consumer and retail firm studies. *American Journal of Agricultural Economics,* 89, 1282-1287.

Piskurich, K., Nelson, D., and Herz, D. (2001). *Maximizing respondent contact in the American Time Use Survey.* Presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Quebec, Canada.

Raabe, S. (2010, June 10). Shopper-card data traced for other uses. *Denver Post.* Available at: http://www.denverpost.com/business/ci_15264783.

Redline, C., Tourangeau, R., Couper, M., Conrad, F., and Ye, C. (2009). *The effects of grouping response options in factual questions with many options.* Presented at the Annual Conference of the Federal Committee on Statistical Methodology. Available at: http://www.fcsm.gov/09papers/Redline_IX-B.pdf.

RTI. (2002). *2001 National Household Survey on Drug Abuse incentive experiment.* Available at: http://www.drugabusestatistics.samhsa.gov/nhsda/methods/incentive.pdf.

Safir, A. (2011). *Survey design and data collection.* Briefing presented to the Consumer Expenditure CNSTAT Panel. Available at: http://www.bls.gov/cex/redpanl1_safir.pdf.

Safir, A., and Goldenberg, K. (2008). Mode effects in a survey of consumer expenditures. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Available at: http://www.amstat.org/sections/srms/Proceedings/.

Schneider, S.J., Gustavson, B., Li, T., Göksel, H., DiGaetano, R., and Edwards, W.S. (2009). *Personal health information management and the design of consumer health information technology: Secondary analysis of data from the Medical Expenditure Panel Survey* (AHRQ Publication No. 09-0090EF). Rockville, MD: Agency for Healthcare Research and Quality.

Schneider, S.J., and Reifer, A. (2011). *Data capture technologies and financial software for collecting consumer expenditure data.* (Report submitted to the Bureau of Labor Statistics). Rockville, MD: Westat.

Singer, E., Gebler, N., Raghunathan, T., Van Hoewyk, J., and McGonagle, K. (1999). The effect of incentives in interviewer-mediated surveys. *Journal of Official Statistics,* 15, 217-230.

Smith, A. (2011). *35% of American adults own a smartphone.* Washington, DC: Pew Internet and American Life Project. Available at: http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_Smartphones.pdf.

SymphonyIRI Group. (2011). *Consumer Network.* Available at: http://www.symphonyiri.com/ProductsSolutions/AllProducts/AllProductsDetail/tabid/159/productid/56/Default.aspx.

Tucker, C. (2011). *Using multiple data sources and methods to improve estimates in surveys.* Discussion presented at the Household Survey Users Workshop. Available at: http://www.bls.gov/cex/hhsrvywrkshp_tucker.pdf.

U.S. Census Bureau. (2011). Section 25. Banking, finance, and insurance. *2011 Statistical Abstract.* Available at: http://www.census.gov/prod/2011pubs/11statab/banking.pdf.

U.S. Department of Agriculture. (2011). *Food assistance and nutrition programs: The National Household Food Acquisition and Purchase Survey (FoodAPS).* Available at: http://www.ers.usda.gov/Briefing/FoodNutritionAssistance/food_aps.htm.

U.S. Department of Energy. (2011). *How does EIA estimate energy consumption and end uses in U.S. homes?* Available at: http://www.eia.gov/consumption/residential/methodology/2009/brief.cfm.

Westat. (2005). *A field test of a multiple diary procedure for the Consumer Expenditure Survey* (Report to the Bureau of Labor Statistics). Rockville, MD: Westat.

Wine, J., and Riccobono, J. (2011). *National Postsecondary Student Aid Study.* Presented at the Household Survey Producers Workshop, Bureau of Labor Statistics. Available at: http://www.bls.gov/cex/hhsrvywrkshp_wine.pdf.

Wohl, J. (2011). *Walmart to give its U.S. store data to Nielsen.* Available at: http://www.reuters.com/article/2011/07/21/us-walmart-nielsen-idUSTRE76K3K420110721.

Yodlee, Inc. (2011). *Yodlee aggregation SDK.* Available at: http://www.yodlee.com/solutions-developer-api-yodlee-aggregation.html.