

Redesign Options for the Consumer Expenditure Survey

Final Report

Authors

David Cantor, Westat

Nancy Mathiowetz, University of Wisconsin-Milwaukee

Sid Schneider, Westat

Brad Edwards, Westat

June 21, 2013

Prepared for:
Bureau of Labor Statistics
2 Massachusetts Avenue, NE
Washington, DC 20212

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

Executive Summary

The Consumer Expenditure Survey (CE) is a program run by the Bureau of Labor Statistics (BLS) to collect data on American consumer spending. The program currently consists of two distinct data collection efforts, the Quarterly Consumer Expenditure Interview (CEI) survey and the Consumer Expenditure Diary (CED) survey, each with its own sample of respondents. Researchers use CE data to study consumer spending in the United States. The data also are crucial to the computation of the Consumer Price Index (CPI).

The data collection methods used in the CE program have not been fundamentally redesigned in several decades. However, evidence strongly suggests that CE data are increasingly yielding biased estimates of consumer spending, especially in certain expenditure categories. BLS is considering a number of approaches to updating and enhancing the CE program, with several goals in mind:

- Improving the quality of CE data by reducing measurement and nonresponse error;
- Diminishing the burden on CE respondents; and
- Containing the costs of the CE program.

The BLS contracted with Westat and the University of Wisconsin – Milwaukee to assist in the planning for the redesign of the CE program. This report is the primary product of that contract. The report has four sections. The first presents background material related to the CE program and the plans to redesign it. The second section proposes changes to the CE data collection methods, and presents the rationale for each proposed change. The third section presents a plan for evaluating the proposed ideas for redesigning the CE methods. The final section presents the estimated costs of conducting the CE program using the proposed, redesigned data collection methods.

Proposed redesign to the CE program. The second section of the report gives the recommendations for the proposed redesign. The main features of the new design include:

- **Use a single sample.** Instead of recruiting one sample that maintains diaries and a separate sample that completes interviews, use a single sample that does both.
- **Reduce the number of interviews.** Instead of conducting 5 interviews at 3-month intervals, conduct 2 interviews, 12 months apart.

- **Reduce the diary data collection period by half.** Ask the respondents to keep expenditure diaries for 1 week, instead of the current 2-week period.
- **Vary the reference period in the interviews.** Instead of using a 3-month reference period for almost all of the expenditure categories in the interviews, vary the reference period as appropriate for the specific expenditures.
- **Reduce the redundancy between the diary and the interview.** Limit the focus of the diary data collection to certain expenditure categories, including food. Use global rather than detailed questions for those expenditure categories in the interview.
- **Create a web version of the diary.** Develop a web-based version of the diary to replace the current paper-and-pencil version. Allow respondents to use their own technology, including desktop and laptop computers and mobile devices, to maintain the diary. Lend a tablet to respondents who need a way to access the diary. Accommodate respondents who insist on using a paper-and-pencil version of the diary.
- **Use the capabilities of the web.** Design the web-based diary with interactive features that help respondents enter their data quickly and easily. Include features that allow respondents to direct questions to the CE staff, and that permit the CE staff to detect respondents who appear to have stopped entering expenditure data.
- **Develop an app for respondents to quickly make note of expenditures.** Develop and offer respondents a smart phone app that permits them to create a record of expenditures quickly at the time of purchase.
- **Have multiple respondents for the diary.** Instead of enlisting just a single respondent in each household, ask all members of age 15 or older to maintain the 1-week expenditure diary.
- **Emphasize the use of records.** Train the interviewers to encourage respondents to refer to expenditure records, such as receipts and bank and credit card statements. Call respondents a few days following each interview to collect any data that the respondents have found on such records.
- **Provide incentives.** Offer the respondents monetary incentives, to increase their motivation. Train interviewers to foster a collaborative relationship with respondents.
- **Reduce the burden for reporting of grocery purchases.** Allow respondents to submit their cash register receipts for grocery store or a supermarket purchases instead of entering those purchases, which often comprise many items, into the diary. Ask respondents to write explanatory notes on the receipts where necessary.

Evaluation. The third section of this report presents a plan for developing and evaluating the proposed new data collection approaches. The evaluation plan is intended to be feasible within the time and financial constraints with which the BLS is likely to be working.

The components of the proposed evaluation are as follows:

- **Development of the web-based diary.** The evaluation will help ensure that the web-based diary is easy to use and widely accepted by the respondents. The evaluation methods include an environmental scan, stakeholder interviews and focus groups, prototype testing, and tests in a usability laboratory. The evaluation will involve respondents who use desktop and laptop computers, and who use mobile devices.
- **Content of the interview and diary.** The evaluation will help determine which expenditures should be covered in the interviews and which expenditures should be covered by the diary. Using field tests, the evaluation will explore the feasibility of covering food, alcoholic beverages, personal products, and possibly apparel primarily in the diary.
- **Use of global questions.** Cognitive laboratory research will explore the feasibility of using global questions in the interview. With global questions, respondents provide a single overall estimate of all expenses within a category, or respond to detailed questions only after they have answered a gate question.
- **Length of the reference periods.** Cognitive laboratory research will also explore the feasibility of varying the reference period for various expenditures in the interview. Reference periods of 1, 3, and 6 months may be appropriate, depending upon the nature of the specific expenditure.
- **Use of annotated grocery receipts.** A small-scale feasibility study will examine the merits of allowing respondents to submit a grocery or supermarket receipt, with any necessary notations, in place of entering every item on the receipt into the diary.
- **Encouraging record use.** A series of experiments could help identify the most effective and feasible ways of persuading respondents to use receipts, checking account and credit card data, and other documentation to improve the accuracy and completeness of their expenditure reports.
- **Integrating the interview and diary samples.** A pilot test and a field test will examine the feasibility and potential drawbacks of integrating the current CE interview and diary samples. This field test could also study the use of global questions. The field test could reveal how the proposed redesigned methods affect the number and level of reported expenditures, the time needed to complete the interview, the perceived burden and perhaps respondent-interviewer interactions.
- **Final field test.** A field test involving about 3,000 households could examine multiple redesign proposals, including recruiting multiple diary-keepers in each household, and offering various levels of incentives.

The proposed evaluation plan is not intended to be comprehensive. The proposed testing plan could benefit from more resources to test alternative designs in larger field tests, as well as increasing sample sizes to increase the power of the tests that are proposed.

Costs for the new design. A major constraint in creating the revised design is the requirement that it be cost neutral. The objectives of the final section of this report are as follows:

- To estimate the annual cost of the revised design in its steady state, after implementation;
- To create assumptions based on recent cost data provided by the Census Bureau; and
- To specify the assumptions used for each cost element.

The BLS provided summary cost data from the Census Bureau for conducting the CE in 2011. Total cost was about \$30,000,000. This cost included both the diary and interview components. It did not include any program costs at BLS.

The Census costs were broken out into four categories:

1.	Total field activity tasks	\$6.4 million
2.	Interview only	\$10.7 million
3.	Diary only	\$3.5 million
4.	Overhead	\$9.2 million

The proposed redesigned methods include only two waves of interviewing per household, which is three less than the current design. That change will lower costs. However, in the redesign, the data collection during the two waves require more interviewer effort than any two waves of the current design. The additional effort in the redesign includes the following:

- **First interview.** The interviewers both conduct the first interview and place the diary. Because there is higher burden for the respondent, more interviewer attempts are required to gain cooperation.
- **First interview.** The interviewers track the respondents' diary entries over the week, and call when no expenditures are being entered.
- **First interview.** The interviewers telephone diary-compliant respondents to close out the diary.
- **First interview.** Interviewers call to obtain data that respondents found on any records or documentation that they located after the interview.
- **Second interview.** Interviewers conduct the second interview in-person.

- **Second interview.** Interviewers call to obtain data that respondents found on any records or documentation that they located after the interview.

This section of the report presents cost calculations relevant to the proposed, redesigned data collection methods. The calculations suggest that the redesign costs are approximately the same as the current design costs. The estimate for the redesign costs is very sensitive to the degree of field effort required to achieve the desired number of completed interviews and diaries and the response rate targets.

Table of Contents

Chapter	Page
I. Introduction: Motivation for Redesign, Goals, and Constraints	1
Motivation for Redesign Effort	1
Goals of the Redesign Effort	3
Constraints	5
II. Design Features	6
Overview	6
Integrated Interview and Diary Sample	7
Panel Design: Number and Frequency of Interviews	11
Content of Diary and Interview Survey	12
Interview: Content and Reference Periods	13
Interview: Global Questions	17
Use of Records	18
Diary Field Period	21
The Use of Technology for the Diary	22
Respondent's Technology for the Diary	31
Active Monitoring of Diary Data Entry	34
Use of Mementos for the Diary	35
Multiple Diarists Per Household	36
Use of Annotated Grocery Receipts	37
Incentives	39
Sample Sizes	42
Length of First Interview	43
III. Evaluation Plan	42
Philosophy	42
Development of Web-Based Diary	42
The Content of the Interview and the Diary	46
The Use of Global Questions	47
Length of Reference Period	47
Use of Annotated Grocery Receipts	48
Record Use: Understanding Record Organization and Willingness to Access	49
Development and Testing of an Integrated Sample	51
Field Test: Global Questions, Diary Content and Reference Periods	52
Final Field Test	53
Statistical Power	55

Summary of Schedule and Costs of Development and Testing Plan	56
Scope of the Research and Development Program	58
IV. Costs for New Design	59
Objectives	59
Assumptions and Limitations	60
Extrapolation of Essentially Variable Costs from Census Cost Information and Westat Experience.....	61
Benchmarking and Sensitivity	66
References	68

<u>Table</u>	<u>Page</u>
1. Proposed reference periods for OPLC categories on the recall interview	14
2. Recurring expenses proposed for 6-month reference periods and variation within reference period	16
3. Proposed sample sizes compared to current design.....	43
4. Power for difference in proportions by sample size, level of rate and percentage point difference between the groups*	56
5. Power for difference of means for selected expenditures, completed diaries per group and effect size*	56
6. Power for difference of means for selected expenditures, completed interviews per group and effect size*	56
7. Summary of expenses for R&D work (\$1,000).....	58
8. Cost elements shared between current design and proposed design.....	62
9. Field representative labor: Contact attempts per consumer unit.....	63
10. Estimated Census interviewing costs in redesign	66
11. Total costs for redesign in steady state at Census.....	66

<u>Figure</u>	<u>Page</u>
1. Data collection plan: Redesigned CE program	10
2. An early screen. An early screen asks respondents what they need to do.	25
3. First screen for entering expenditure data. The respondent first enters a description of the purchase, and clicks a radio button to indicate the category of the purchase. The questions that come next are dependent upon the category that the respondent selects.....	26

<u>Figure</u>		<u>Page</u>
4.	Next screen for entering expenditure data. If the respondent entered an expenditure that fell into the category “any other product or service,” the next screen would ask for the price without tax, and the date of the purchase.	27
5.	Later screen for entering expenditure data. This screen would appear for all expenditures except those in the category “food and drink away from home.”	28
6.	Help screen. If a respondent clicked the link “help” for the question about the cost of an expenditure, the screen that popped up would offer brief overall guidance, and list common sources of confusion about the price. When the respondent clicked an item on the list, another screen would pop up to explain how to handle that situation.	29
7.	Timeline for redesign tasks.....	57

Introduction: Motivation for Redesign, Goals, and Constraints



The Consumer Expenditure Survey (CE) is a program run by the Bureau of Labor Statistics (BLS) to collect data on American consumer spending. The CE program consists of two distinct data collection efforts, the Quarterly Consumer Expenditure Interview (CEI) survey and the Consumer Expenditure Diary (CED) survey.

In both the CEI and the CED, one household member serves as the respondent for the household economic unit, called the consumer unit (CU). The CEI and CED use separate samples of CUs. In the CEI, respondents are interviewed once per quarter for five consecutive quarters. In the initial interview, respondents report their demographic and family characteristics, their ownership of major durable goods, and their CU's expenditures for the past month. The initial interview serves as a bounding interview; the expenditure data are never analyzed. In the subsequent four interviews, the respondents report on their CU's expenditures for the past 3 months. Those interviews are generally conducted in person, but increasingly they must be conducted by telephone.

In the CED, respondents record all household expenditures on diaries over two consecutive 1-week periods. The diaries are in a paper-and-pencil format. Before completing the diaries, the CUs provide data about their demographic and family characteristics, much like the CEI CUs do.

Estimation of total expenditures is based on an integrated approach, drawing upon data from both the CEI and CED. Researchers use data from the CE program to study purchasing behavior in the United States. The data are also used for calculating the Consumer Price Index (CPI; Bureau of Labor Statistics, 2009).

Motivation for Redesign Effort

BLS has decided to redesign the methods of the CE program. The central goal of the redesign is to improve the quality of CE data while controlling the cost of the program. Comparisons between estimates based on CE data and estimates based on data from the Personal Consumption Expenditures (PCE) program of the Bureau of Economic Analysis point to deterioration in the

quality of CE data, especially over the past decade (e.g., Slesnick, 2001).¹ Several sources of measurement error may be affecting CE data. For example, respondents interviewed by telephone may be more likely than those interviewed in person to under-report expenditures (Safir and Goldenberg, 2008; McGrath, 2005). As more respondents have to be interviewed by telephone, the quality of CEI data may be declining. In the CED, the respondents' expenditure reports tend to diminish over the course of the 2-week diary keeping period (Silberstein and Scott, 1991). That finding suggests that the quality of CED data is affected by the length of the reporting period.

Nonresponse error may also be contributing to the disparity between CE and PCE data. The period of increasing discrepancies between the CE and the PCE estimates has also been a period of fluctuating CE response rates (Goldenberg, McGrath, and Tan, 2009; Safir, 2011). More important than the nonresponse rate is nonresponse error--the strength of the association between response propensity and the expenditures being measured. Chopova et al. (2008) assessed nonresponse bias in the CEI and found differential participation as a function of race, income and dwelling unit characteristics as well as evidence of nonresponse bias for particular expenditure categories.

Both the CEI and CED data collection efforts are burdensome--the interview with respect to the cognitive burden of retrospective recall for 3-month reference period, the diary with respect to the sheer recording burden when respondents must use paper-and-pencil forms. The quality of CE data might improve if the burden on the respondents could be diminished.

The quality of CE data might also improve if a CU could have multiple respondents, instead of just one. Grootaert (1986) found that expenditures, especially for personal items such as clothing, footwear, and services, were significantly higher in households that had multiple respondents, as compared with households that were limited to a single respondent. Similarly, a field study conducted by BLS in 2006 (Edgar et al., 2006) found that the use of multiple diaries per CU resulted in an increase number of expenditure items reported as well as an increased dollar value of the expenditures.²

¹ Similar to Slesnick (2001) one has to also question the extent to which the external data used to benchmark CE data accurately represent consumer expenditures

² The experimental study had a lower response rate than the production data and completed interviews required more trips on the part of field interviewers as compared to the production cases. In addition, the CED experimental study also yielded comments from some CU members that they felt the individual diaries were redundant and unnecessary since the unit makes purchases collectively.

Technology may facilitate the collection of information from multiple members of the household. In particular, the task of keeping a diary might be less burdensome if the respondents could use an electronic version of the diary.

The separate samples in CEI and CED may also contribute to problems with CE data. With these separate samples, a very large number of CUs must report data about topics such as demographic background and ownership of durable goods. Integrating the two samples may diminish the overall burden on the public by reducing the number of CUs that have to provide this information. The bounded interview design, in which data from the first interview are not used for estimates, is expensive and places burden on the CUs. Eliminating the bounding interview may help contain the cost of redesigning the CE program.

The CUs in the CEI and CED receive no incentive, monetary or otherwise. Adding an incentive might improve the quality of CE data.

In summary, the motivation to redesign the CE program stems from a desire to meet several goals:

- Improve the quality of CE data by reducing measurement and nonresponse error;
- Diminish the burden on CE respondents; and
- Contain the costs of the CE program.

In order to reach these goals, BLS is conducting the Gemini Project, a comprehensive effort to redesign the CE program. The Gemini Project is considering changes to many aspects of the CE program. The purpose of this project was to recommend a specific design and research program that addresses each of the above goals.

Goals of the Redesign Effort

A redesign of the Consumer Expenditure Surveys has not been undertaken since the late 1970s. Since that time, technology, the retail environment, and the CE program itself have radically changed. The Gemini Project offers an opportunity to modernize the CE design, incorporating the best data collection practices while addressing the unique requirements of the CE program. The BLS web site (<http://www.bls.gov/cex/geminiproject.htm>) explains the Gemini Project goals:

The mission of the Gemini Project is to redesign the Consumer Expenditure surveys (CE) in order to improve data quality through a verifiable reduction in measurement error, with a particular focus on under-reporting. The effort to reduce measurement error is to be explored in a manner consistent with combating further declines in response rates. Accordingly, any expected benefits of survey design changes will be balanced against any potential negative effects on response rates. The primary objective of the Gemini Project is to develop a detailed research planning document for a redesign of the CE surveys. The planning document will describe the priorities, individual steps, timeframe, resource needs, and costs required for the development, pilot testing, evaluation, and implementation of a redesigned CE survey or surveys. To allow for an unpredictable budgetary environment, the planning document will address both a complete redesign of the CE surveys as well as more limited modifications to the current design. A secondary objective of the project is to generate a planned research agenda that will guide the development and implementation of CE research studies throughout the Gemini project lifecycle.

In this report, we present our recommendations for redesigning the CE program. We also present our recommendations for evaluation studies, and our analyses of the costs involved.

We have attempted to approach the redesign and the evaluation studies with a concern for both measurement and nonresponse error; that is, our recommended redesign attempts to improve data quality while reducing, or at least not increasing, nonresponse error. Inherent in our approach is a philosophy that the respondents should understand that the interviewer's goal is not only to achieve a completed interview but also to foster a collaboration between the respondent and the interviewer. Most respondents currently are ill-informed as to the nature of their data reporting tasks—in part because interviewers are reluctant to request the respondents' commitment for fear of harming response rates. We believe that the success of our recommended redesign will require clear communication to the respondent about the nature of the data reporting tasks and the goals of accurate and complete data. Our approach incorporates methods such as asking the respondents to use expenditure records, preparing the respondents for the initial interview via the advance letter, and encouraging respondents to complete their diaries daily. We also believe that the interviewers' task requires evaluation across a range of metrics beyond response rates, including the rate of record use among respondents, the rate of diary compliance, and the assessment of data quality (controlling, of course, for respondent characteristics). Best practices with respect to monitoring of field interviewers are moving toward the incorporation of quality control measures often used in centralized telephone interviewing—for example, computer assisted recording of interviews (CARI),

so as to measure and provide feedback to field interviewers across a broader range of metrics than merely response rates (Cantor, 2010).

We recognize that this philosophy underlying our design represents to some extent a paradigm shift in how the CE has been fielded in the past. However, this total quality approach to field efforts is the direction in which the field is moving and is not dissimilar to the approaches increasingly used in other federal data collection efforts.

Constraints

The Gemini project – and the design outlined in this report – face several constraints. Most notable are time limitations. The redesign must be completed, including evaluation, in 5 years. In addition, the redesign faces two limitations related to costs: (1) the redesign efforts and evaluation must not exceed \$1 million per year (total \$5 million) and (2) the redesigned survey must be cost neutral. It is with consideration of these constraints that we have outlined both a set of design features for the new CE as well as a plan to evaluate those design features.

We also note that a major constraint is the lack of sufficient empirical knowledge at this point in time to specify a definitive approach to the CE data collection. Many of our recommendations outlined in Section II of this report lack detail because the details are dependent upon the outcome of the evaluation projects outlined in Section III. The recommended design should be viewed as the framing of a house, the details of which will depend upon the results of the evaluation studies. In an ideal world, we would have the opportunity to return to the design recommendations in five years, when the empirical data from the evaluation studies are available. In addition, details of evaluation field studies—particularly the larger field studies—will be informed by earlier, more exploratory efforts (e.g., laboratory-based studies, focus groups); as such, the details for some of these studies are not included and cannot be outlined at this time.



Overview

Our recommendations for redesigning the CE program include a number of significant departures from the current design:

- **Use a single sample.** Instead of recruiting one sample that maintains diaries and a separate sample that completes interviews, we recommend recruiting a single sample that does both.
- **Reduce the number of interviews.** Instead of conducting 5 interviews at 3-month intervals, we recommend conducting just 2 interviews, 12 months apart.
- **Vary the reference period in the interviews.** Instead of using a 3-month reference period for almost all of the expenditure categories in the interviews, we recommend varying the reference period according to the nature of the expenditure.
- **Reduce the redundancy between the diary and the interview.** We recommend limiting the focus of the diary data collection to certain expenditure categories, including food. We recommend using global rather than detailed questions for those expenditure categories in the interview.
- **Reduce the diary data collection period by half.** We recommend asking respondents to keep expenditure diaries for 1 week, instead of the current 2-week period.
- **Create a web version of the diary.** Currently, the diary is available only in paper-and-pencil format. We recommend developing a web version. Respondents can then use their own technology or use a government-supplied tablet device to access the diary.
- **Use the capabilities of the web.** We recommend designing the web-based diary with interactive features that help respondents enter their data quickly and accurately, especially for complex expenditures. We also recommend including features in the web-based diary that allow respondents to direct questions to the CE staff, and that permit the CE staff to detect respondents who appear to have stopped providing data.
- **Develop an app for respondents to quickly make note of expenditures.** We recommend developing a smart phone app that permits respondents to create a record of an expenditure at the time of purchase.

- **Include all members of the CUs.** Instead of enlisting just a single respondent in each CU, we recommend asking all members of age 15 or older to complete the 1-week expenditure diary and participate in the two interviews.
- **Reduce the burden for reporting of grocery purchases.** Currently respondents must enter each item that they purchased at a grocery store or a supermarket into the diary. We recommend allowing the respondents to submit their cash register receipts instead. The respondents would write notes on the receipts where necessary.
- **Emphasize the use of records.** We recommend encouraging respondents to refer to expenditure records, such as receipts and bank statements, so that they can provide accurate data.
- **Provide incentives.** We recommend increasing the respondents' motivation by offering them monetary incentives.

In this chapter, we present our recommendations and explain their rationale. We describe why we believe that our recommendations may improve data quality, foster respondent compliance, and control data collection costs while meeting the objectives of the CE program.

Integrated Interview and Diary Sample

Recommendation: We recommend adopting a single sample design that incorporates two waves of in-person interview data collection conducted 12 months apart, with a single week of diary-keeping immediately following the first wave of CE interview data collection.

The redesign of the CE program must improve the quality of the data while also meeting certain data requirements (e.g., for the CPI). This challenge needs to be addressed without additional burden to the CU or cost to the overall CE budget. For that reason, we recommend keeping both interview and diary methods of data collection. The interview is most suited to certain expenditure categories such as appliances, while the diary is most suited to other expenditure categories such as apparel. However, we recommend discarding the current design in which separate samples participate in the CEI and CED, and replacing it with a single sample design. With our recommended design, a single sample would participate in two in-person interviews, 12 months apart, and keep an expenditure diary for one week after the first interview.

Outside CE data users clearly desire a single sample design that integrates diary and interview data collection methods (Edgar, 2005; Henderson et al., 2011). Moving to a single, integrated design

provides flexibility in index construction for the CPI (Casey, 2010). It offers micro-data users the advantage of accessing a single data source for the best estimate of expenditures and for studying change over time.

Successfully recruiting respondents is a major task facing any survey, especially when the survey entails substantial burden. An integrated design with a single sample reduces the costs related to recruiting two independent samples. Also, with an integrated design, the interviewers can establish rapport with the respondents during the interview data collection; the respondents may then be motivated to provide good-quality data on the diary.

Other countries have moved to an integrated design for their consumer expenditure surveys (Horsfield, 2010; Tremblay, Lynch, and Dubreuil, 2011). Both Canada and the United Kingdom have an integrated design. However, the design that we propose for the United States differs in several fundamental ways from the Canadian and the UK designs:

- The Canadian and UK designs require that a CU maintain a diary for 2 weeks after the initial interview. We recommend a 1-week diary following the interview.
- In the Canadian and UK designs, respondents record data in their diaries for all expenditures, with a few exceptions such as utilities and mortgages. We recommend minimizing redundancy between the interview and the diary, so as to focus the diary specifically on those expenditures that are best collected as they occur. This approach may reduce burden and improve the quality of the expenditure reports. The Canadian and UK designs employ cross-sectional samples, in which a CU is included for a single interview and a 2-week diary reporting period. Our proposed design – albeit increasing burden from those of the Canadian and UK designs – attempts to leverage a tradeoff among recruitment, burden, data quality, costs, and user requirements with two interviews 12 months apart but a single request for a 1-week diary.
- The Canadian experience does not include the use of an incentive. The UK design does include an incentive for the diary paid at the individual level but not for the interview. We recommend incentives for both the diary and the interview.
- In the Canadian design, respondents are permitted to provide receipts, instead of making diary entries, for restaurant and grocery expenditures. The UK design does not have this feature. We recommend that respondents be allowed to provide receipts in lieu of diary entries for expenditures at grocery stores and supermarkets.

We are keenly aware of the response rates that Canada and the UK achieved with their integrated designs. The Canadian experience, implemented in 2010, achieved interview response rates around 66 percent during its first 2 years, with less than 70 percent of those who completed the interview completing the diary (effective response rate of approximately 46% for the diary). We could not find

separate estimates for the response rates in the interview and the diary in the UK, but the overall response rate for the Living Costs and Food Survey is 51 percent.

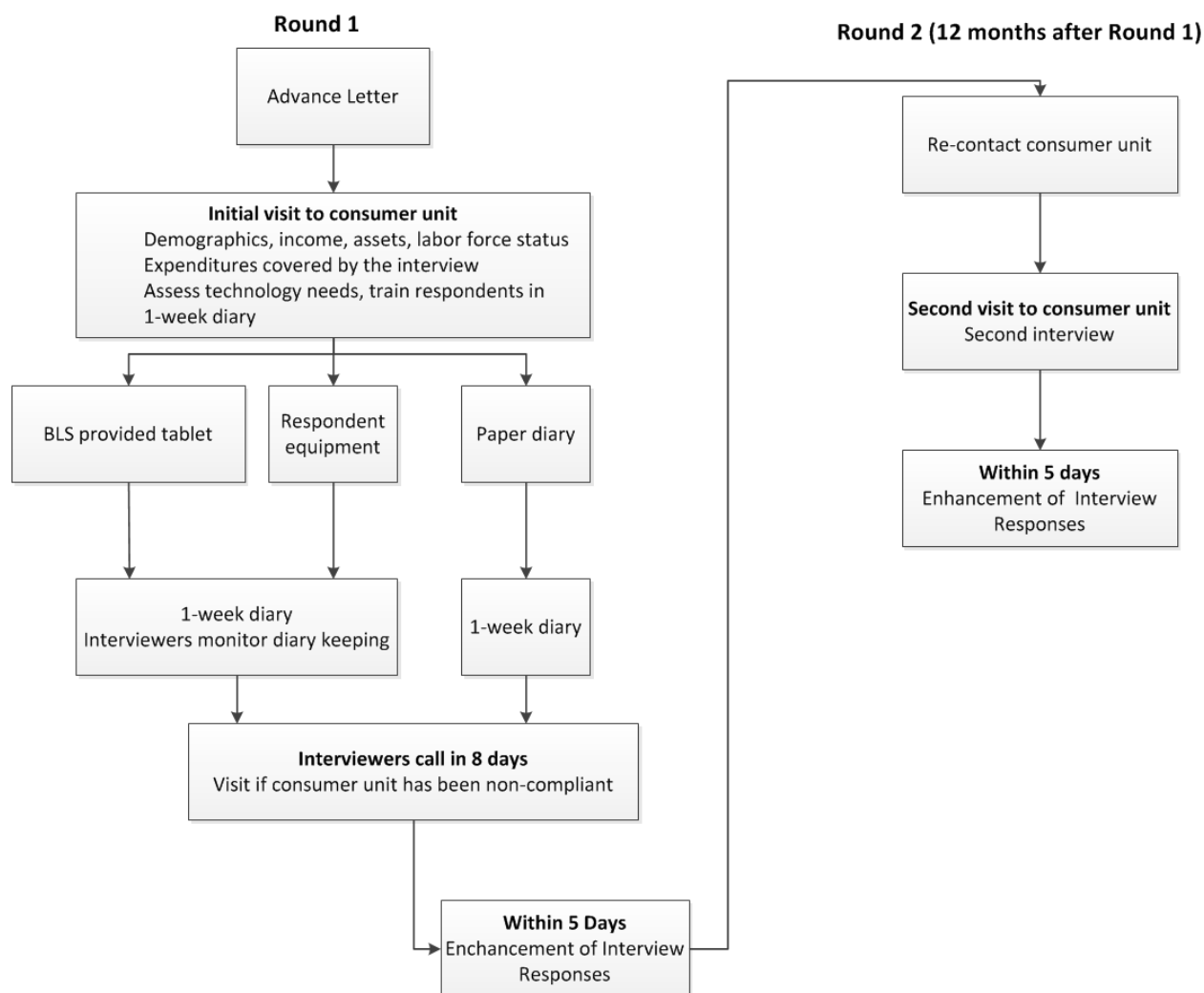
Clearly the Canadian interview response rate is lower than the stand-alone CEI response rate in the United States. It is unclear (and unknowable without testing) whether the discrepancy is due to the difference in the nature of the request (interview plus diary) or a function of differing interviewing staffs in the two countries.

A possible field protocol for an integrated interview-diary sample might be as follows:

1. Advance letter. The advance letter encourages CUs to collect their expenditure records. It informs them of the incentive.
2. Initial face-to-face contact with sampled CU. During the initial visit, the interviewer conducts the first interview, and places diary materials with the CU.
3. Followup visit. The interviewer returns to the CU 8 days following the start date of the diary data collection. During this visit, the interviewer collects receipts for grocery and supermarket expenditures. In some cases, this face-to-face visit will not be necessary; the interviewer and CU could interact through telephone calls and U.S. Mail. The interviewer pays the incentives.
4. Follow-up contact to enhance interview data. The interviewer contacts the household to revise data from the initial interview using any records and other information that the respondent has found since that interview.
5. Advance contact for the second interview. The interviewer telephones, emails, or sends U.S. Mail to the CU 7 days prior to second interview. In this contact, the interviewer reminds CU members to gather records needed for the interview.
6. Second interview. The interviewer visits the CU and conducts the second interview, 12 months after the first interview. The interviewer pays the incentives.
7. Follow-up contact to enhance the second interview. The interviewer contacts the household to revise data from the initial interview using any records and other information that the respondent has found since that interview.

This data collection plan is shown graphically in Figure 1.

Figure 1. Data collection plan: Redesigned CE program



With respect to respondent burden, this design has two advantages. First, it reduces the number of interviews from five to two. Second, there is a reduction in the redundant collection of expenditure items. Currently all items are collected on both the interview and the diary. The recommended design restricts content on the diary. Relative to the current CED, the burden on a single CU goes up with the requirement of completing the interviews. But when viewed in the context of the current CEI, where five interviews are required, the tradeoff is a net reduction. The question that needs to be addressed is whether the response rates and the quality of the expenditure reports on the diary suffer after integrating the diary with the interview.

Panel Design: Number and Frequency of Interviews

Recommendation: Limit participation for a sampled CU to two waves, conducted 12 months apart.

The current CE interview schedule consists of five quarterly interviews, with the first interview serving as a bounding interview. Data from the bounding interview are not utilized in estimation; therefore 20 percent of the collected data are not used. The cost of collecting data that are never used for estimation is large, both in terms of interviewing costs and respondent burden. In addition, the repeated interviewing leads to inefficiencies with respect to variance estimation (Westat, 2011). Our design recommendations are driven by the priorities established for CPI estimation. To maximize the quality of data, minimize costs, while achieving the precision needed for the CPI, we recommend a two-wave panel, with interviews conducted 12 months apart. By limiting the panel to two interviews, our proposed design reduces burden on the CUs. Prior research suggests that reducing the number of interviews used for estimation from four to two, while shifting the reference period for some expenditures, may also significantly improve the precision of the estimates (Westat, 2011).

A major disadvantage of our recommended 2-interview design is the loss of 12 months of consecutive data for a CU. Although the design permits annual comparisons – from a quarter in year t to a quarter for year $t+1$ – the importance of the loss of annual data should not be underestimated, especially for data users outside of BLS. The current five-interview design does provide micro-data users the ability to model full year consumer expenditure patterns. Although the five-interview panel data are not a requirement for the CPI, they are desirable for analytic uses of the data.

Content of Diary and Interview Survey

Recommendation: We recommend including only food, alcohol, apparel, and personal product expenditures in the diary data collection, and including all other expenditures in the interview data collection. The interview should also include global estimates for the expenditures covered by the diary.

We reviewed the source (CEI or CED) for the current CE estimates for the Universal Classification Codes (UCCs) of each of the 125 Office of Prices and Living Conditions (OPLC) categories. Based on this review, we developed an initial proposal for splitting the UCCs between the interview and diary. The OPLC categories with 85 percent of expenditures from UCCs taken from a single source were assigned to that source.³ For the CED, these are food, alcohol, and personal care products. For the interview, these are recurring expenditures (e.g., utility bills), large items (e.g., cars, appliances) and intermediate items that can't be reliably collected on the diary (e.g., floor coverings).

We reviewed each of the remaining OPLC categories with two criteria in mind. The first was whether the items could be logically assigned to diary data collection. The expenditure items to be included on the diary should be salient and memorable. Food and alcohol expenditures are examples. Respondents buy food on a frequent basis and the universe of eligible events is easily remembered. Personal care products don't necessarily fit in with food or alcohol, but they are small, frequently purchased items that are sold at many of the same stores that sell food. Other types of expenses which are not as salient or as frequently purchased, such as "tools, outdoor equipment and supplies," should not be assigned to diary data collection. If they were included, respondents could not easily define and remember the universe to be included on the diary.

"Apparel," however, could be assigned to diary data collection. "Apparel" comprises several different OPLC categories (men's, women's, children's, footwear, infants', and toddlers'). While these items are not purchased as frequently or at the same stores as food, alcohol or personal care products, they do constitute a clear, cognitively salient, grouping that respondents could remember.

The second criteria determining the split between expenditures to be collected on the diary and the interview is the precision of the estimate. If the items within an OPLC category are not purchased very often, relying on the diary will not yield an estimate with adequate precision. This is the case for

³ An exception to this is "other personal services," which is assigned to the interview.

some of the apparel OPLC categories. Boy's apparel, for example, has about 66 percent of expenses currently taken from the interview survey. About one-third of expenses for the other apparel OPLC categories come from the interview.

We propose that the diary cover expenditures for food, alcohol, personal care products and apparel. All other expenditures will be covered on the two recall interviews. If there are certain types of apparel that cannot be estimated with enough precision from the diary, then the interview could include selected apparel items.

For expenditure items covered by the diary the interview should include global rather than detailed questions.

Interview: Content and Reference Periods

Recommendation: The content of the interview survey will vary across the first and second interview. The first interview will include all expenditures proposed for the interview. The second interview will include the same expenditures, except it will not include the largest expenditures.

Our recommended design calls for two interviews at each CU, spaced 12 months apart. In the current CEI, the reference period for all expenditure items is the same: 3 months. With our recommended design, the reference periods for the expenditure items will vary between 1, 3, and 6 months.⁴ The 6-month items will be covered only in the first interview, while the 1 and 3 month items will be covered in both interviews. This procedure will minimize respondent burden while still providing adequate data. Table 1 provides the reference periods that we recommend for each OPLC category.

Assignment to particular reference periods was based on examining the salience and regularity of expenditures in each category. Salient purchases, such as large expenses (cars, major household appliances), were assigned to a 6-month period. Very regular expenses, such as rent and utilities, were also assigned to a 6-month period. Most other expenses were assigned to a 3-month period.

⁴ At the proof of concept presentation we recommended a 12-month period. After review of the sample size requirements and the concern with burden, we have shortened the reference period to 6 months.

The exceptions were those OPLC categories that had over half of their UCC expenses taken from the diary for the current CE estimates. These were assigned to a 1-month period.

The 6-month items are divided into two types – recurring and non-recurring. Recurring items are typically regular bills or expenses that repeat over the reference period. Non-recurring items are occasional purchases; for the most part, these purchases are of relatively expensive items. For each item, respondents report the amount of the expenditure and the month it was incurred. At first glance, this reporting could be a very burdensome task for the recurring expenditures that involve 6 separate occurrences of a monthly expenditure. However, the size of many of the recurring expenses should be constant across the reference period and should be relatively easy to report. Table 2 provides a classification of the recurring expenditures, indicating whether or not each expenditure is likely to have a constant amount over the 6-month period.

Table 1. Proposed reference periods for OPLC categories on the recall interview

Expense category	Reference period
<i>Regular expense</i>	
Mortgage interest and charges	6
Owners' equivalent rent of primary residence	6
Owners' equivalent rent of vacation/second residence	6
Personal property taxes	6
Property taxes	6
Electricity	6
Fuel oil and other fuels	6
Internet services	6
Natural gas	6
Telephone services	6
Water, sewer, and trash collection services	6
Household insurance	6
Long-term care insurance*	6
Commercial health insurance	6
Adult care	6
Medicare payments	6
Child care	6
College tuition	6
Other school tuition tutoring and test preparation	6
Rent of primary residence	6
Medicare prescription drug premiums*	6
CE life and other personal insurance	6
CE vehicle finance charges	6
Trade-in allowance, new motorcycles*	6
Amount motorcycle sold or reimbursed*	6
Funeral expenses	6

Table 1. Proposed reference periods for OPLC categories on the recall interview (Continued)

Expense category	Reference period
<i>Large expense</i>	
Furniture	6
Amount vehicle sold or reimbursed	6
Boats, RVs, and other recreational vehicles	6
New cars and trucks*	6
Used cars and trucks*	6
Motorcycles	6
CE finance charges excluding mortgage and vehicle	6
Catered affairs	6
Other housing expenses	3
Maintenance and repair supplies	3
Appliances	3
Information and information processing other than Internet services	3
Unpriced maintenance and repair services	3
Floor coverings	3
Motor vehicle parts and equipment	3
Other apparel services	3
Leased and rented vehicles	3
Trade-in allowance, leased vehicles	3
Legal fees and financial services	3
Trade-in allowance, new vehicles	3
Motor fuel	3
Motor vehicle fees	3
Motor vehicle maintenance and repair	3
Dental services	3
Hospital and related services	3
Other medical services	3
Physician services	3
Airline fares	3
Lodging away from home	3
Rent as pay	3
Meals as pay	3
Household operations	3
Medical equipment and supplies	3
Food on out-of-town trips*	3
School meals*	3
Food prepared by CU on out-of-town trips*	3
Occupational expenses*	3
CE deductions for social security (3 month)	3
CE payroll deductions to retirement plans (3 month)	3
CE non-payroll deposit to retirement plans	3
CE Alimony	3
CE support for college students	3
CE cash contributions to charities, education, and other organizations	3
CE other cash and non-cash gifts	3
CE assets and liabilities	3

Table 1. Proposed reference periods for OPLC categories on the recall interview (Continued)

Expense category	Reference period
51% to 84% of UCC expenses from CE interview	
Information and information processing other than Internet services	3
Photography	3
Sporting goods	3
Video and audio	3
Household textiles	3
Jewelry and watches	3
Educational books, and supplies	3
Movie, theater, sporting event, and other admissions	3
Recreation services	3
Recreational reading materials	3
Other public transportation	3
Pets, pet products, and services	3
Tobacco and smoking products	3
51% to 84% of UCC expenses from diary	
Tools, hardware, outdoor equipment, and supplies	1
Household supplies*	1
Other household equipment and furnishings	1
Sewing	1
Medicinal drugs	1
Miscellaneous	1
Other recreational goods	1
Postage and delivery services	1
Motor vehicle insurance	1
Other personnel services+	1

* New categories.

+ 100 percent of these expenses are sourced from the diary for Current CE estimates.

Table 2. Recurring expenses proposed for 6-month reference periods and variation within reference period

Expense category	Same amount each time unit?
Mortgage interest and charges	Yes
Owners' equivalent rent of primary residence	Yes
Owners' equivalent rent of vacation/second residence	Yes
Personal property taxes	Yes
Property taxes	Yes
Electricity	No
Fuel oil and other fuels	No
Internet services	Yes
Natural gas	No
Telephone services	Yes
Water	No
Sewer	No
Trash collection services	Yes
Household insurance	Yes
Long-term care insurance	Yes
Commercial health insurance	Yes

Table 2. Recurring expenses proposed for 6-month reference periods and variation within reference period (Continued)

Expense category	Same amount each time unit?
Adult care	Yes
Medicare payments	Yes
Child care	Yes
College tuition	Yes
Other school tuition tutoring and test preparation	No
Rent of primary residence	Yes
Medicare prescription drug premiums	Yes
CE life and other personal insurance	Yes
CE vehicle finance charges	Yes

Interview: Global Questions

Recommendation: We recommend using global questions in the interview when appropriate.

We recommend using global questions wherever possible in the interview. This will minimize respondent burden while still generating adequate data. Rather than asking respondents to report on a number of specific expenditures, the interviewer would ask about entire classes of expenditures. Respondents would then have a way to report that they had no expenditures for a class of items (e.g., men's clothing), rather than having to respond to questions about each item within each class, as the current interview requires.

Global questions can take one of two forms. With the first form, respondents provide a single estimate for a broad class of expenditures. The empirical evidence on the accuracy of the data obtained using this strategy relative to using more specific questions is mixed (e.g., BLS, 2010). However, the evidence seems strong that the accuracy of the data depends on the specific expenditure items involved. Beatty (2010) and Creech et al. (2012) found that global questions yielded comparable or superior data for expenditures in homogenous categories that are amenable to respondent cognitive organization.

It could be argued that global questions are most appropriate to collect data about regular purchases, because respondents tend to overestimate these expenditures when they try to decompose them into specific classes. However, Beatty (2010) argued that global questions are appropriate for eliciting

data about highly frequent, relatively simple behavior (e.g., getting a drink from a fountain), but not about many of the expenditures covered on the CEI. Clearly more research is needed into this issue.

The second type of global question serves as a filter question, which allows respondents to skip out of entire sections of the interview if they respond by saying that they have made no purchases in a particular category. As Kreuter et al. (2011) have found, however, this type of filter question may encourage respondent's "satisficing," once they realize they can avoid follow-up questions if they answer "no."

The natural unit for global questions is the OPLC categories that will be used for analysis. In our evaluation plan (in Section III, below), we recommend testing global questions to determine which of the OPLC categories are suited to this methodology. We recommend that the expenditure items that are covered by the diary be among those that are covered in the interview using global questions.

Use of Records

Recommendation: Encourage respondents to refer to records for the interview. First, the advance material should ask the respondents to gather records that they could use during the first interview. Interviewers should contact the CUs to ask them to gather records before the second interview, too.

Currently, CE interviewers mention the benefits of recordkeeping to respondents in the CED and the CEI. The CE paper diaries have a pocket in the back for the respondents to store receipts and other records. During the interviews, respondents can refer to any readily-available records. We recommend that the interviewers go even further to help ensure that respondents gather and use their expenditure records. Item #3 in the Consensus Design Document of May 7, 2012 supports the development of techniques to enhance the use of records.

The rationale for asking respondents to refer to records is simply that records are likely to contain accurate information. Respondents who try to recollect their expenditures without referring to records are more likely to forget some expenditures and commit errors reporting the ones that they

remember. In other federal data collection efforts, such as the Medical Expenditure Survey (MEPS), respondents are strongly encouraged to save their records.

Edgar (2010) investigated how often respondents actually referred to records during the CEI. She found that almost 39 percent of the respondents never or almost never referred to records, and about 19 percent only occasionally referred to records. She also found that the use of records was associated with improved data quality, particularly for certain types of expenditures, such as payments for utilities and health insurance.

Recently, two contractors conducted field tests to investigate the feasibility of encouraging respondents to keep and refer to expenditure records. RTI International researchers (Gelsen et al., 2011) found that respondents with certain demographic characteristics, such as non-Hispanic whites and homeowners, were more likely than others to refer to records. Respondents were especially likely to have records for recurrent expenditures, higher-cost expenditures, and for recent purchases. Respondents tended not to print out receipts for online purchases. The researchers found that the use of records improved the quality of the respondents' reported data; however, the respondents themselves placed confidence in their own recollections even when they did not refer to records.

NORC researchers (Sjoblom and Lee, 2012) found that CE respondents generally complied with requests to keep expenditure records. Overall, respondents could produce records corresponding to about 80 percent of all reported expenditures. Respondents in the lowest income groups and living in one-person households were among the least likely to keep records. Respondents tended not to have records for certain types of expenditures, mostly because no receipt or other record was ever created for these expenditures, which included public transportation fares, and concession stand and vending machine purchases.

The NORC researchers reported that over 80 percent of the respondents thought that keeping records was easy, and over 65 percent reported that they were able to keep all of their relevant expenditure records. Nearly half reported that they could easily persuade other members of the household to provide records. Over 80 percent stated that they did not feel uncomfortable sharing all of their expenditure records. Over 94 percent used a special receipt holder that the interviewers provided to them. The researchers examined the records that the respondents provided and found that very few contained any personally identifiable information like the purchaser's name.

The RTI and NORC field tests suggest that further encouraging recordkeeping may be relatively easy to implement and may help improve data quality. However, the impact will not affect all respondents or all expenditure categories equally.

In Section III (below), we propose an evaluation to identify the most effective methods to encourage respondents to gather and use their expenditure records.

Recommendation: Prior to starting an interview, the interviewer should ask respondents to gather summary records (e.g., credit card statements, checking account statements) that might assist in recall of expenditures.

Prior to the start of each interview, the interviewers should ask the respondents to access credit card statements, bank statements, checkbooks, or other records that may help the respondents recall their purchases over the reference period. Banks keep a database of all checks that the respondents have written. These databases always contain the amount of each check and the date that it cleared. Sometimes they also contain the payee, at least for some of the checks. The databases may also contain a digital image of each check. Also, credit card companies keep a database of all charges, including the date, payee, and dollar amount. The respondents could view these databases online, or download and print out the contents of the databases. They could also see the same data on their checking and credit card account statements, which they may receive in the U.S. Mail or view online. At least by one estimate, approximately 70% of all consumer purchases are contained in electronic records like this (Westat, 2011).

Recommendation: At the conclusion of each interview, interviewers should ask respondents to find expenditure records over the next several days. The interviewer should call the main household respondent back after a pre-specified period of time (see below) and ask for any updates or changes to information as a result of new information contained in the records retrieved since the interview.

We recommend the interviewers ask the respondent to gather expenditure records during the days after the interview is completed. The interviewer would ask the respondent to collect any records that were not on hand during the interview, with a special emphasis on the larger purchases reported using the 6 or 3 month reference period. These records would represent expenditure information for items that were reported from memory at the first interview. Or it may be expenses that were completely left out during the first interview. The interviewer would call the respondent to collect

this updated information. After the first interview, this callback would occur within five days after the diary is collected. At the second interview, the callback would occur within 5 days after the interview. This “enhancement” could take one of two forms. One would have the interviewer pull up interview data that had not come from records and ask the respondent if a record had been found. If so, the information would be updated using the new information. A second possible format would be open-ended, with the interviewer asking the respondent to produce any records they might have gathered. For each record found, the interviewer would locate the section of the interview that needed updating.

For purposes of budgeting (Chapter 4), we have assumed this is conducted as a followup telephone interview. Alternatively, the respondents could enter the information on a special online form. In Section III (below), we propose an evaluation to study the use of records this way.

Recommendation: Interviewers should encourage the use of records when completing the diary.

The respondents should verify their diary entries against receipts, online purchasing records, checking account statements, and credit card statements. Interviewers could furnish the respondents with special receipt holders, and with stickers reading “Remember to get a receipt!” for their wallets.

Below we also recommend that respondents be permitted to provide grocery store and supermarket receipts, with appropriate annotation, in lieu of entering data into the diary. This procedure would spare the respondent a significant amount of burdensome and possibly error-prone diary entry.

Diary Field Period

Recommendation: Limit the diary data collection to 7 days.

The recommendation to move to a single week rather than a two-week diary period is based on the empirical literature that demonstrates a decline in reporting as the length of the diary period increases. The literature on this issue with respect to CE diaries is relatively limited. The most salient findings include these:

- For daily or weekly expenditure diaries, several authors have reported systematic declines in expenditure estimates. They observed that the estimates of the first week are

higher than those of the subsequent week or weeks, and that the average estimate of the first day of the first week is greater than the overall average estimate (Cho, Eltinge, and Steinberg, 2004; Dippo et al, 1977; Pearl, 1977; Silberstein and Scott, 1991).

- The mean expenditure amount drops on average 20 percent from the first to second day of the first week and 10 percent from the first to the second week (Silberstein and Scott, 1991).
- Regardless of the week, weekly expenditure means decline as a function of the amount that recall, rather than records, is used. Expenditures had the highest means when there is no recall, followed by when there is partial recall and the lowest when all expenditures are provided by recall.

Other studies incorporating diaries have seen similar fall off. Examples include health diaries (Verbrugge, 1980), food diaries, and time use diaries.

This research suggests that CE data quality may improve if the diary-keeping period were shortened from 2 weeks to 1 week. Moreover, respondents may perceive less burden and be more likely to cooperate when their diary keeping task lasts only 1 week.

The Use of Technology for the Diary

Recommendation: Create a web version of the diary. Design the web-based diary with interactive features that facilitate data entry.

We concur with item #5 in the Consensus Design Document of May 7, 2012, calling for the development of a web-based diary. The rationale for our recommendation to develop a web-based diary is straightforward:

- Research in other data collection projects suggests that replacing paper-based data collection methods with electronic methods may potentially improve data quality;
- Many respondents will be reluctant to download software for the purpose of data entry, but few will feel uncomfortable accessing a web-based diary;
- The current CE paper diary provides a relatively cumbersome data entry process; and
- A web-based diary can have capabilities which facilitate data entry and which are infeasible with paper diaries.

Research from other data collection projects suggests that converting from paper-based diaries to electronic diaries may improve the quality of the data that the respondents provide (Iida et al., 2012; Krishnamurthy, 2008; Stone et al., 2002). Those projects include psychological research in which respondents were asked to record their activities or self-observations. Those projects also include clinical trials of investigational medications in which patients recorded events such as side effects. The results of those projects suggest that respondents tend to enter their data more promptly and accurately on electronic diaries, as compared to paper diaries. The CE program might similarly benefit from a transition to an electronic diary.

An electronic diary can be created two ways: The diary can run on the web, or it can run as software on smart phones, tablets, and laptop and desktop computers. With a web-based diary, respondents use a browser to access the diary and enter their data. With a software-based diary, respondents run the diary software on their own computers or mobile devices; after they enter their data, the software uploads the data to a central database.

We recommend that the BLS create a web-based diary, not a software-based diary, because many respondents are likely to be reluctant to download the software. Only 43 percent of American adults who own a smart phone ever download apps onto it (Smith, 2012). Michael Link (personal communication, September, 2012) found that respondents were often unwilling to download software for data collection in Nielsen projects, even though those same respondents sometimes download games and other software that they know little about. Some respondents may view the requirement to download software as an unnecessary burden. Some respondents may encounter difficulty trying to download the software. Some may even suspect that the software could be dangerous.

By contrast, respondents are not likely to have misgivings about accessing a web-based diary. Respondents would not need to download any software to access the web-based diary. They would access it in the same manner that they access any web site.

The current paper diaries are burdensome in a number of ways. Respondents must locate the appropriate day and category on the diary forms before entering each expenditure. All the instructions are in a single, lengthy section at the start of the diary, where they are not easily available when the respondent needs them. Respondents using the paper diary are therefore likely to guess about how to handle complicated expenditures or ambiguous purchase prices. In addition, respondents using the paper diary have limited ability to interact with CE staff. A web-based diary need not have those drawbacks (Davis, Stinson, and To, 2002).

A well-designed web-based diary would enable respondents to enter their expenditure data with a minimum of effort. BLS has designed and will soon field test a web-based version of the current CE paper diary. The design of that web-based diary generally conforms to the design of the paper diary; the respondent enters expenditure data onto online tables that resemble the forms in the paper diary.

We recommend a different design for the web-based diary. With our recommended design, the web-based diary prompts the respondents to enter their expenditures interactively by answering questions, rather than by entering expenditure data directly into tables. Respondents would not need to navigate to specific pages based on the date or the category of the purchase.

An example of one possible version of our proposed design, the respondents would first indicate what they wished to do. They would see a screen like the one in Figure 2. They would indicate whether they needed to enter an expenditure, view or change information about an expenditure that they already entered, or enter information about an expenditure that they already entered but then returned or exchanged.

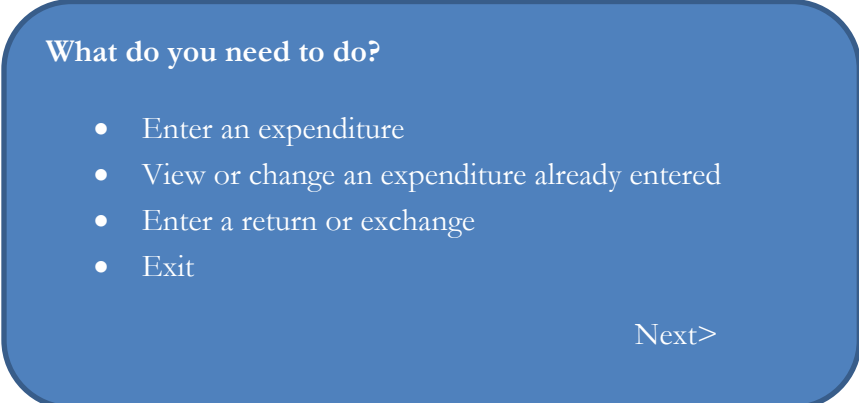
Respondents who wanted to enter data about an expenditure would next see a screen that prompted them to enter a description of the expenditure and to select the category that best described the expenditure (Figure 3). The screen that the respondents saw next would depend on the selected category.

If a respondent was entering data about an expenditure in the “other” category—that is, not food away from home, nor food for home consumption, nor apparel—then the next screen would ask the respondent to enter the amount paid, and the date of the expenditure (Figure 4). If the respondent had entered an expenditure in the “food and drink away from home” category, the screen would ask for the price with tax and tip, and the screen that came next would have the questions “meal type,” “where purchased,” and the questions about the kind and cost of alcoholic drinks that may have been purchased with the meal. Similarly, if the respondent had entered an expenditure in the “clothing, shoes, jewelry, and accessories” category or the “food and drink for home consumption” category, the screens would ask for the specific information required by the CE program.

On the succeeding screen, the respondent would indicate whether the expenditure was for someone not on the household list (Figure 5). The screen would display the household list, in case the respondent did not remember the list.

The respondents would see the word “help” below each question. When they clicked the “help” link, they would receive guidance for that particular question. For example, Figure 6 shows a screen that respondents might see if they clicked the “help” link below the question about the dollar amount of an expenditure. The “help” screen provides general instructions, along with a list of possible sources of confusion, such as purchases in which the respondent receives a discount for the merchandise, or pays for the merchandise in installments. When the respondent clicked on the text describing the situation, a text box would pop up providing guidance about how the respondent should compute the price. With this “help” feature, the respondent will feel that adequate guidance is always quickly available; as a result, respondents may be less tempted to guess how to enter the data.

Figure 2. An early screen. An early screen asks respondents what they need to do.



What do you need to do?

- Enter an expenditure
- View or change an expenditure already entered
- Enter a return or exchange
- Exit

Next>

Figure 3. First screen for entering expenditure data. The respondent first enters a description of the purchase, and clicks a radio button to indicate the category of the purchase. The questions that come next are dependent upon the category that the respondent selects.

What did you buy or pay for?

[Help](#)

What kind of purchase was this?

- ☐ Food or drink away from home
- ☐ Food or drink for home consumption
- ☐ Clothing, shoes, jewelry, or accessories
- ☐ Any other product or service

[Help](#)

[<Back](#) [Next>](#)

Figure 4. Next screen for entering expenditure data. If the respondent entered an expenditure that fell into the category “any other product or service,” the next screen would ask for the price without tax, and the date of the purchase.

[The text that the respondent typed to describe the expenditure, such as “Hammer” or “Plumber’s fee” appears here]

What was the total cost, without tax?

\$.

[Help](#)

What was the day of the purchase?

- ☐ Sunday, April 1
- ☐ Monday, April 2
- ☐ Tuesday, April 3
- ☐ Wednesday, April 4
- ☐ Thursday, April 5
- ☐ Friday, April 6
- ☐ Saturday, April 7

[Help](#)

<Back Next>

Figure 5. Later screen for entering expenditure data. This screen would appear for all expenditures except those in the category “food and drink away from home.”

[The text that the respondent typed to describe the expenditure, such as “Hammer” or “Plumber’s fee” appears here]

Was this purchase for someone in your household?

- ☐ Yes, it was for someone who is in my household
- ☐ No, it was for someone who is NOT in my household

[Help](#)

Your household

John Q. Doe
Jane Roe-Doe
Billy Doe
Sally Doe

<Back Next>

Figure 6. Help screen. If a respondent clicked the link “help” for the question about the cost of an expenditure, the screen that popped up would offer brief overall guidance, and list common sources of confusion about the price. When the respondent clicked an item on the list, another screen would pop up to explain how to handle that situation.

Enter the total cost, but do not include any tax or tip. So if you paid \$10.00 for a haircut and added a \$2.00 tip, enter “\$10.00.” If you bought a hammer and paid \$5.00 plus 40 cents tax, enter “\$5.00.”

Sometimes the cost is not completely clear. Click the situation that applies to this purchase for more explanation:

I paid for something now, but will not use it until later. For example, I bought a train ticket today but will use it next week.

I bought something and received it, but paid nothing now. I will not pay until later.

I used a coupon to reduce the price of the item.

I bought something but made only a partial payment now. I'll pay the rest later.

I received a discount that applies to the total purchase price, not to the individual items that I bought.

I will receive a refund of all or part of the price later. For example, I will send in a coupon to receive a rebate.

I received a “buy one get one free,” or “buy one, get another at half price,” or “buy one, get something else free,” or similar discount.

Close

Respondents might need to view or change data about the expenditures that they have already entered, or to report that they returned or exchanged an item that they already entered as an expenditure. In those cases, the web-based diary might present a list of the expenditures that the respondent had already entered. The list might include the description of the expenditure, the price, and the date. If the respondent clicked on one of the expenditures on the list, the next screen would show all the entered data about the expenditure. The respondent could then delete the expenditure, or change any of the data. The respondent could indicate that the purchase was returned and the payment was refunded. Or, the respondent could indicate that the purchased item was exchanged; the next screen would ask the respondent about the merchandise received in exchange.

The system could provide some level of error checking. For example, it could ask respondents to verify apparent errors, such as entering \$250 as the price of apples, or categorizing a shirt as food.

The web-based diary could include a way to interact with CE staff. For example, respondents might click a link to create and send a message to the staff. When respondents used the data entry system later, they would read the reply. Also, the web-based diary might even include “live chat,” like many e-commerce sites today. With “live chat,” respondents would interact with CE staff in real time.

We emphasize that many approaches are feasible for designing an interactive diary. We have presented here only one of many possibilities. Moreover, some respondents may prefer to enter their data into online tables rather than with an interactive series of questions. A well-designed web-based diary will be flexible, allowing respondents to enter data in the way they prefer.

We recommend that the interviewers spend the necessary time to explain the web-based diary to the respondents. In addition, we recommend that the interviewers call the respondents 2 days into the diary period. The interviewers can then address any difficulties that the respondents may have encountered; Horsfield (2010) used that procedure in a different expenditure data project.

Our proposed solution offers an adequate level of data security. The web-based diary would use a secure Hypertext Transfer Protocol Secure (HTTPS) protocol, like the web sites of most banks and credit card companies use. All respondents would be required to enter a user identification and password before they entered expenditure data. After they finished entering data, they would log off of the system. If they did not log off, the session would time out.

Respondent's Technology for the Diary

Recommendation: Allow respondents to use their own technology to enter their expenditure data on the web-based diary. Provide tablets to respondents when necessary. Offer the paper diary to respondents who prefer to provide data that way.

We recommend allowing respondents to use their own smart phones, tablets, laptop computers, or desktop computers to enter their expenditure data on the web-based diary. Research from the Pew Internet and American Life project suggests that a very large proportion of American adults have Internet access conveniently available:

- About 80 percent of American adults have access to the Internet, although poorer and less educated individuals are disproportionately less likely to have Internet access (Zickuhr and Smith, 2012).
- About 62 percent of American adults have broadband Internet access at home (Zickuhr and Smith, 2012).
- About 25 percent of American adults have a tablet computer (Rainie, 2012).

Those findings suggest that most CE respondents could be expected to have convenient access to the Internet and be well accustomed to using web sites for a variety of purposes. They may expect to be permitted to enter their data using their own technology.

There are clearly challenges in developing a web-based diary that respondents can use on a range of platforms. The term “Responsive Web Design” is defined in Wikipedia as “an approach to web design in which a site is crafted to provide an optimal viewing experience—easy reading and navigation with a minimum of resizing, panning, and scrolling—across a wide range of devices, from desktop computer monitors to mobile phones.” BLS would need to apply the techniques of Responsive Web Design in order to make a suitable web-based diary for the CE program.

When a browser requests a web page from a server, it provides certain information, such as what kind of browser it is, and the hardware and operating system it is using. With effective Responsive Web Design, the server configures the web page according to that information.

Developers today have at their disposal numerous Responsive Web Design tools including Hypertext Markup Language Revision 5 (HTML5), Cascading Style Sheets Level 3 (CSS3), JQuery

Mobile (www.jquerymobile.com), and Sencha Touch (www.sencha.com). Developers use these tools to create web sites that accommodate consumers who use a vast diversity of computers and mobile devices.

Some providers of web-based surveys use Responsive Web Design tools to ensure that all respondents can complete their surveys, no matter what technology they own. For example, Voxco claims that web surveys developed with its Acuity4 system can be completed on almost any platform (<http://www.voxco.com/en/product/web/online-survey-software.php>). Other web survey systems, including REDCap and Vovici, make the same claim. Clearly data collection using Responsive Web Design tools is becoming prevalent.

A web-based diary for the CE program need not contain extensive graphics, or lengthy text, or a very large number of pages. A skilled designer will be able to create the web-based diary so that it ran on the range of contemporary platforms consistently and usably.

CE interviewers, during the initial visit to a CU, could ask respondents about the devices that they owned and their willingness to use those devices for data entry. The interviewers could ensure that the devices were suitable for using the web-based diary. If not, the respondent would be asked to use a tablet supplied by the government.

We do not concur with the National Academy of Sciences (NAS) panel's recommendation that BLS provide all CUs with a tablet as the sole data entry method, and not allow respondents to use their own computers and mobile devices (Dillman and House, 2013). We recognize that the NAS recommendation would simplify the data collection process in some ways. However, we also believe that the tablet-only approach has drawbacks:

- If BLS provided a tablet to every CU, BLS would need considerable resources to replace the tablets if significantly superior technology became available. Our recommendation allows more flexibility to BLS to adjust to changing technology over the next 5 to 10 years, when the CE redesign will be implemented.
- Respondents would have no way to enter data when they are away from the tablet, such as at work or out shopping or on trips. When respondents must wait until they return to the tablet before they can report their data, the delay is likely to diminish the quality of the data that the respondents report (Krishnamurthy, 2008; Stone et al., 2002);
- Many respondents are likely to feel burdened and inconvenienced if they cannot use their own, readily-available computers and mobile devices to report their data. Some respondents may not trust a government-supplied computer.

We nonetheless do recommend lending tablets to respondents who wish to use the web-based diary but do not have the technology to do so. Some respondents may wish to enter their expenditure data on the web-based diary but lack adequate access to the Internet. For example, some respondents may have Internet access only at work. Some respondents may have access to the Internet only with a smart phone, but not wish to enter data on its tiny screen. The CE interviewers would lend these respondents a tablet computer. Because a respondent's home may not have WiFi, the tablet would have cellular connectivity with the Internet. The tablet would be locked down so that respondents could use it only to enter expenditure data on the web-based diary. More than one respondent in the CU could use the tablet. The respondents would not be restricted to using the tablet as their sole data entry device; they could also enter data by accessing the web-based diary with any other device that they had the opportunity to use, such as a desktop computer at work.

Some respondents may not wish to use any technology at all. These respondents would enter their expenditure data using paper forms.

Below we recommend allowing each CU to have multiple diary respondents instead of only one. Within a single CU, some respondents might enter data using their own devices, while other respondents used a tablet provided by the interviewer, and other respondents relied on paper forms.

Respondents who use the BLS-issued tablet would not have access to any data other than their own. Even if they were sharing the tablet with others in their CU, they would be able to view only their own data after entering a username and password.

Our recommended solution—allowing respondents to use their own technology but lending tablets to respondents who lack technology—is not a new idea. Other data collection efforts have successfully employed similar strategies. For example, the firm Knowledge Networks, now a unit of GfK, fielded web surveys to a nationwide panel of households. Most of the household members accessed the surveys using their own computers. Knowledge Networks gave a netbook to about 30 percent of the households, however, because they did not have their own technology to access the Internet (<http://www.knowledgenetworks.com/knpanel/index.html>).

We considered the possibility that the interviewers could lend respondents smart phones as well as tablets. However, it is unlikely that many respondents who lack Internet access will wish to enter their data using a smart phone. Many of the respondents who do not have Internet access may have limited Internet experience (Zickuhr and Smith, 2012). They might well prefer the larger screen of a tablet.

We recommend that the respondents who receive a tablet use the browser to enter their data over the Internet. However, an alternative approach is also possible: The diary could be pre-loaded onto the tablet. Respondents could enter their expenditure data directly on the tablet. The data would then be automatically uploaded to a central database. That approach might minimize data charges. That approach also has a disadvantage: updating the diary software may be difficult when a new version is developed.

Active Monitoring of Diary Data Entry

Recommendation: Take advantage of the web-based diary design to monitor and encourage daily diary entries for all CU members.

Shifting the diary to a web-based instrument allows for real time monitoring of each CU's entries into the diary. We recommend that the instructions to the respondents outline the need for daily data entry, even if there have not been any expenditures.

The diary would be designed so that respondents could indicate "no expenditures today." When respondents indicated that they had no expenditures, the diary could present a checklist of potentially overlooked expenditures such as "vending machine purchases" and "automatic utility payments." This checklist could help respondents recall expenditures that they had forgotten. Respondents could check one of the expenditures on the list; they would then be asked to enter the relevant CE data. When respondents checked "No expenditures today," they would be affirming that they did not neglect to enter expenditures for the day, but actually had no expenditures to report. CE staff would then know that the respondent had not simply decided to stop reporting CE data. If a respondent did fail to interact with the data collection system at all for a period of several days, then the interviewer could call, text, or email the respondent. In Section III below, we suggest evaluating the effectiveness of this procedure in improving data quality.

Use of Mementos for the Diary

Recommendation: Develop a smartphone app for respondents to make note of expenditures quickly.

Some respondents may wish to create mementos when they purchase something, to help them remember to enter data about that purchase. For example, respondents rushing through a checkout line at a store might take a photograph of the merchandise that they were buying. Respondents who wished to use mementos would download an app onto their smart phones. The app would permit them to take a photograph, record a voice memo, or type or write a short note. The app would automatically upload the memento, encrypted, to a central database. The next time that the respondents accessed the web-based diary, they would see the photograph, read the note, or hear the voice memo, reminding them to enter data about the purchase.

Respondents would not be required to enter a user identification or password to create a memento on their smart phones. Respondents would typically be in a rush when they created a memento. They may have just handed a payment to a busy cashier, for example, and not have time to enter a user identification and password into a smart phone. For that reason, the proposed app would allow them to take a photo, record an audio memo, or type or write a note very quickly. The memento app could help some respondents remember their expenditures. Of course, only respondents who were willing to download this app could use it. Pew research suggests that smart phone owners often like to create mementos. For example, about 82 percent of American adults who own a smart phone use it to take photos (Duggan and Rainie, 2012).

Some respondents might choose instead to use their own apps to create photos, audio memos, or typed and written notes. Since these mementos would not be uploaded, the respondents would need to remember to check them when entering data into the diary.

Multiple Diarists Per Household

Recommendation: We recommend that all CU members, ages 15 and older, complete a 1-week diary.

A limited body of empirical research suggests that switching from the current single respondent per CU design to a multiple respondent per CU design may improve the quality of the data collected with the CE diary, primarily by reducing under-reporting. Examining household expenditures in Hong Kong between 1979 and 1980, Grootaert (1986, p. 943) found that switching from the single diary to a multiple diary approach lead to a 10.7 percent increase in expenditures for a 2-week diary period. The expenditures most affected were for clothing and footwear, transportation and vehicles, services, and miscellaneous goods. Even among nuclear families, there was a significant impact related to the use of multiple diaries, with significant increases for clothing and footwear as well as services expenditures.

Similarly, a field study conducted by BLS (Edgar et al. 2006) found that the use of multiple diaries per CU resulted in an increased number of expenditure items reported as well as an increased dollar value of the expenditures. Specifically, the BLS researchers found increases in the mean number of reports ranging from 5.2 percent for “other expenditures” to 6.8 percent for food at home. Looking at mean expenditures, the researchers reported significant increases in expenditures at the CU level, on average \$281 higher in week 1 for CUs assigned to the individual diaries and \$540 higher in week 2.

Grootaert’s (1986) research suggests that full participation by all household diary-keepers is not the key to increasing the reporting of expenditures. Rather it is the request to the household to keep multiple diaries that leads to the reporting of increased expenditures. A related advantage of this approach is that it could significantly reduce the perceived burden of the survey.

The field protocol for the Living Costs and Food Survey (UK) consists of a household-based Computer-Assisted Personal Interview followed by a request that every individual aged 16 and older keep a detailed record of expenditures for a 14-day period. Children ages 7 to 15 are also asked to keep a simplified diary of daily expenditures. According to the Office of National Statistics (2011), about 86 percent of the children ages 7 to 15 in responding households maintained their own diaries. Incentives were paid post completion of the diaries: £10 for adults and £5 for children. The impact

of the request for multiple diaries on response rates cannot be ascertained as no relevant controlled research has been published; current response rates average around 50 percent for the survey.

A field test that compared single and multiple diary data collection protocols was conducted in 2011 as part of the Household Food Acquisition and Purchase Survey (Mathematica, 2011). With the multi-diary protocol, all adults (ages 19+) and youth (ages 11-18) in the households kept diaries of all household food acquisition over a 7-day period. The field test found that households assigned to the single diary condition reported more spending and consumption of food at home while households assigned to the multiple diary condition reported more food consumption away from home. As a result of the field test, three kinds of diaries were adopted in 2012: (1) a Primary Respondent Book; (2) Adult Food Book for adults age 19 and older other than the primary respondent; and (3) Youth Food Book for those age 11-18.⁵

Other examples of data collection efforts that involve a request for multiple diarists within the household include the National Household Travel Survey (U.S. Department of Transportation, 2011), time-dairy studies (e.g., Fu, et al., 2002) and the nutritional diary study associated with the National Health and Nutrition Examination Survey (NHANES).

The details of how to place multiple diaries could be an issue, should this recommendation be adopted. We recommend that the person who knows the most about the expenditures in the household receive instructions on the diary and this individual inform the other diarists of their duties.

Use of Annotated Grocery Receipts

Recommendation: Allow respondents to provide annotated receipts for grocery or supermarket expenditures, in lieu of making diary entries.

Research by NORC (Sjoblom and Lee, 2012) suggests that respondents are willing to save receipts, especially receipts for apparel and food purchases (either for consumption at home or away from the home), and that most receipts can be coded without further input from the CE respondents. In

⁵ The field test also included an incentive experiment, comparing a household level incentive of \$50 vs. \$100. The \$100 incentive resulted in response rates nine percentage points higher than the \$50 level. The final field protocol incorporated the use of multiple diaries, incentives at the household level, and multiple telephone contacts during the diary week.

addition, that research identified specific detailed data required by CE that were not available from receipts, including the type of meal and outlet type for food away from home, whether food purchased for consumption at home was fresh, frozen, bottled, canned, or some other form, and the gender and age of the person who would wear purchased clothing, shoes, jewelry, and accessories (Sjoblom and Lee, 2012; p31-32).

The burden on CE respondents would be reduced if they could provide receipts rather than make diary entries, especially for grocery and supermarket purchases, which often include a large number of items. The challenge, however, is to have records that meet the demands of the CE data requirements, and provide sufficient detail to code the expenditures into OPLC categories. The revision of the OPLC categories, eliminating many detailed distinctions, greatly facilitates using grocery store receipts in place of diary entries. For example, the category “fish and seafood” no longer requires distinguishing among canned, fresh, and frozen fish or seafood.

Some stores put UPC codes on their receipts (for example, Wal-Mart and Target). With the UPC code, a product’s OPLC category is easily identified. However, other stores do not put UPC codes on their receipts. Respondents who purchase groceries at those stores will most likely need to annotate their receipts to permit coding by either the interviewer or a central coding operation. Under our redesign proposal, there is also an opportunity for interviewers to follow-up with respondents for receipt information that is not clear from their original annotations.

If replacing diary entries with receipts proves to be successful for food and apparel expenditures, the procedure could be extended to other expenditure types, such as apparel and personal care items. In Section III, below, we propose an evaluation to study how successfully respondents could provide receipts in place of making diary entries.

Incentives

Recommendation: Incorporate incentives at all stages of the CE data collection to both improve response rates and motivate reporting. Include an incentive in the advance letter, and provide incentives to the CE interview respondents and to every diarist in the CU.

Although incentives have most often been considered as a means to boost response rates, we recommend incentives as a means to control total survey error. That is, we view incentives as a tool not only to improve the response rate, but also to motivate respondents to use records for the interview and to record diary entries on a daily basis. We believe incentives are a key to the success of our proposed one-sample design. With appropriate incentives, respondents may be more motivated to provide complete and accurate data for the interview and the diary.

The timing of incentives, the use of CU and person-level incentives, and the amount of the incentive should be designed to achieve several goals:

- Reduce nonresponse;
- Reduce the level of effort to achieve specific response rates;
- Motivate respondents to gather and use records for the interview; and
- Motivate daily recording of expenditures by each respondent.

The survey literature lacks a unifying theory by which to guide the use, timing, and level of incentives. Leverage saliency theory suggests that incentives should be most effective when response propensity is low (Singer, et al., 1999); here we would expect a survey like the CE interview or diary would benefit from an incentive due to the low topic saliency for most CUs. Several authors (e.g., Baumgartner and Rathbun; 1997; Berlin, et al., 1992; Groves, Singer, and Corning, 2000) have found differential impacts of incentives as a function of topic salience, although the compensating effect of incentives has not been consistent in the literature. What the literature fails to address is how incentives impact decision heuristics related to whether or not to participate or the level of effort respondents are willing to exert to complete the survey task.

Two meta analyses (Church, 1993; Singer, 2002) have served as the foundation for three tenets considering incentive use in surveys for the past decade: (1) prepaid incentives are more effective than promised incentives; (2) monetary incentives improve response rates as compared to

non-monetary incentives; and (3) the effect of incremental increases in incentive amounts vary as a function of subject matter and population subgroup.

Evidence of diminished returns associated with prepaid incentives has motivated some commercial organizations (e.g., Arbitron, Nielsen) to experiment with promised incentives. O'Hare (2008) found that promised incentives (\$5, \$10, and \$10 gift cards) all significantly improved radio listening diary return rates, with the \$10 promised incentive raising the return rate by 11 percentage points. Some organizations mix prepaid and promised incentives. For example, in a second study reported by O'Hare (2008), a prepaid \$1 incentive included in an advance letter significantly improved willingness to place diaries in the household and a \$5 "thank you" sent following agreement to complete the diary boosted diary return rates significantly.

The monetary incentives used in past research range from token incentives of \$1 to \$3 to far more substantial amounts. Many of the more substantial incentives are linked to face-to-face interviews that represent significant respondent burden, in terms of time, detail of information, or sensitivity of the information. For example, the Health and Retirement Survey currently offers a prepaid incentive of \$50 to \$80 per married couple. When the incentives are substantial, they are typically paid at the time of the interview is completed. Examples include a \$40 incentive for the National Survey of Family Growth, a \$50 incentive per round for the Medical Expenditure Panel Survey (MEPS) for a total of \$250 for those households that complete all five rounds, and a \$90 incentive for the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The NESARC incentive is split – \$45 up front to the respondent, \$45 at the completion of the interview. Also, the National Health and Nutrition Examination Survey (NHANES) provides varying levels of incentives at the individual level, with the levels based on age and willingness to participate in a medical exam at the time of the interview. For example, those ages 16 and older who agree to the physical exam are offered a \$125 incentive at the time of the interview; those under the age of 12 with no physical exam are offered a \$40 incentive. In addition, NHANES provides other examination-based incentives (e.g., \$40 for wearing a physical activity monitor).

The survey methodology literature is relatively rich in demonstrating the effects of incentives on response rates, level of effort, and data quality, including research specifically focused on the CE interview and diary. As illustrative of the type of study often conducted, we report here an experiment involving MEPS (Hicks, et al., 2011). MEPS is a panel survey in which households are interviewed five times over 2.5 years, collecting information about all medical care utilization and expenditures, with each interview lasting on average 60 to 90 minutes. Three levels of incentives were tested: \$30, \$50, and \$70. Response rates were highly associated with incentive level, with rates

of 58.4 percent, 66.7 percent, and 71.1 percent, respectively, in the first round (all comparisons statistically significant at $p < .01$). In addition to response rates, the study investigated the effects of incentives on level of effort and several proxy measures of data quality. With respect to level of effort, the study found that in general, the higher the incentive, the lower the average number of contacts to complete a case. With respect to data quality (permission to access medical records; rates of missing data on selected items; proportion of households reporting at least one physician visit), the study reported no difference across incentive groups.

The effect of incentives has been studied with respect to the CEI and the CED. McGrath, et al. (2007) reported the results of a controlled incentive experiment involving the diary, comparing no incentive, a \$20 incentive, and a \$40 incentive per CU. The incentives were distributed to the CUs with the mailed advance letter, prior to diary placement. The response rate did not vary across the three experimental groups, but substantive differences were detected. McGrath, et al. (2007, p.36) stated in their conclusions:

Respondents who received an incentive reported both a higher number of expenditures, as well as higher levels of expenditures. For total spending across the 2-week Diary period, mean reported spending for the incentive cases was about \$60 higher, whereas median spending for incentive cases was about 10 percent higher than the control group (\$110). As expected, the \$40 group reported more spending than the \$20 group. Of the four main spending categories, reported spending increased the most in the 'Other' category and least for 'Food at Home'. CUs who received the incentive also reported more purchased items when compared to the control group. Those in the \$20 incentive group reported an average of 3.5 more items than the control (an increase of 5.4 percent), while the \$40 incentive group reported approximately 5.9 more item (an increase of 9.1 percent).

The authors were concerned with measurement bias – that is, that the higher spending reported by the incentivized CUs was, in part, caused by the respondents' ability to spend the incentive itself. The data did not permit clearly delineating between increased reporting due to motivational effects of the incentive and increased reporting due to increased income. The recommendation from the study was to not incorporate incentives for the CU diary sample, in part, due to the lack of significant effects with respect to the response rates.

Goldenberg, McGrath, and Tan (2009) report the findings from an experiment involving the use of prepaid incentives (debit cards) included with advance letters for the CE interview. Two levels of incentives were tested: \$20 and \$40. Both were sent via priority mail. These two incentive

conditions were compared to two no-incentive control conditions: an advance letter sent via first class mail and an advance letter sent via priority mail. Incentives were paid only in advance of the first interview (in a 5-interview panel design). The \$40 prepaid incentive led to higher response rates in the first interview (as compared to the regular mail control), in the second interview (compared to all the other conditions) and in the fourth interview (compared to the two no-incentive conditions) but there were no differences in response rates for the third and fifth interview. (Remember, the incentive was offered only in the advance letter prior to the first interview). Further analysis by Goldenberg, et al (2009) suggests that the reduction in nonresponse was due specifically to a reduction in noncontact rates. Similar to other studies (Berlin, et al, 1992; RTI, 2002; Hicks, et al., 2011), the CE incentive experiment found that the \$40 incentive reduced level of effort, in this case, number of personal visit contact attempts, as compared to the control group or no incentive priority mail group. In all these cases, the incentives at least paid for themselves by these reductions in effort.

With respect to data quality, Goldenberg et al. (2009, p. 5997) summarized their findings as follows:

...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 Incentive \$20 group on 7 of 11 measures, but some of the differences were very small and most were not statistically significant.

To date, however, no incentive has been incorporated in the CE interview program. We recommend including incentives in the redesigned CE program.

Sample Sizes

Recommendation: We recommend including 14,000 CUs per year in the CE program.

Table 3 provides the sample sizes that we recommend for each annual sample. Using our recommendation, the CE program would collect comparable amounts of data for each of the expenditure categories as it collects using the current design.

The current design collects 12 months of data over four interviews for 7,000 households, yielding 84,000 household-months (12 x 7,000). Our proposed design collects 6 months of data for 14,000

households, yielding the same number of household-months ($6 \times 14,000$ for the 6 month reference period; $2 \times 3 \times 14,000$ for the 3 month reference period). The exception is for the items with a 1-month reference period that we recommend for the interview survey. In the case of those items, there are two interviews for 14,000 households, yielding 28,000 household-months. However, the items that we recommend for a 1-month reference period are small purchases that have been estimated from the diary, which currently yields fewer collection months.

Table 3. Proposed sample sizes compared to current design

	Current design	Proposed design
CUs interviewed	7,000	14,000
Interviews completed	35,000	28,000
Bounding interview	7,000	—
Data interview	28,000	28,000
Diary		
CUs	7,000	14,000
Diary weeks	14,000	14,000

A similar calculation can be completed for the diary collection. In the current design, there are 7,000 households in the CED sample, with each CU yielding 2 weeks of data for a total of 14,000 household-weeks. Under the proposed design, there are 14,000 CUs in the sample, yielding 1 week of data for the same total number of household-weeks.

If it is desirable to collect an additional week's worth of diary data, then CU's could be asked to complete the diary at the second interview as well. Of course, this would add both cost and burden to the proposed design.

We note that the current design carries a significant design effect when estimating annual expenditures due to repeating interviews up to four times for the same CU for the interview and two times for the diary. This design effect carries a significant cost for a number of expenditures (Westat, 2011). Our recommended design almost eliminates this design effect. By spacing the interviews 12 months apart, the recommended design significantly reduces the overlap for a particular annual period, but still allows for individual analysis of annual change.

Length of First Interview

Our recommended redesign of the CE program decreases overall burden on a single household in several different ways. First it reduces the total number of interviews from five to two. Second, the

diary data collection period is reduced from two weeks to one week. However, the first visit to the household using our recommended procedure will likely be longer than it is using the current procedures. One reason is that the interview will be asking for a six month reference period for a subset of OPLC categories (Table 1). This will increase the number of people who report an expense in these categories. For regular monthly expenditures (e.g. utility bills), the respondent will be asked to report six separate numbers, rather than three. As shown in Table 2, however, many of the six month expenditures do not change from one month to the next (e.g., mortgage, internet services, rent). Specifically, 19 of the 34 six-month categories should be identical across all months. For the larger expenses that are not regular (e.g. furniture, cars), the frequency of purchase is very low and only a small percentage of respondents will report a purchase even when the reference period is extended to 6 months. Therefore, extending the reference period to 6 months should not increase the interview by more than an average of a couple of minutes.

A second reason why the length of the first interview will increase is the interviewer will spend some time setting up the diary data collection. The interviewer will determine the type of technology the respondents will use, instruct them on the diary task, and explain how to distribute the diary to other members of the household. We estimate these tasks will take approximately 15 minutes. This is based on an estimate that it takes 20 minutes to place the diary in the current design. We have taken off 5 minutes from this based on the observation that the 20 minutes includes interviewers briefing respondents about the CE and its goals. In the proposed design, respondents will already be familiar with the CE and the type of data being requested.

Assuming the interview is approximately 50 to 60 minutes (based on current timings), this will increase the length of the visit to 65 to 75 minutes. While this is longer than the current CEI, this is not excessive relative to other in-person surveys that provide the respondent with an incentive, such as MEPS.



Philosophy

The evaluation plan that follows is a set of laboratory and field studies that include usability tests, feasibility tests, and experiments using controlled designs. Our recommended evaluation plan should not be viewed as an exhaustive set of studies that provide BLS with the best possible empirical data to inform a revised CE design. Rather, we have attempted to identify the minimum set of studies that are absolutely essential for redesigning the CE program in order to stay within the time and cost parameters. To meet the time schedule as well as the budget for evaluation work, some design decisions will unavoidably need to be made without additional empirical data. We believe, for example, that it would be too costly and time consuming to experimentally compare a web-based diary and a paper-and-pencil diary. We therefore assume that the redesign will incorporate a web-based diary, as we recommend, and we do not call for an evaluation of that idea. In the last section of this chapter we comment on the overall level of funding that is being proposed for the redesign.

A critical issue in the evaluation plan is determining what evaluation work can be accomplished while the web-based diary is being developed. Key features of our recommendations (e.g., the one-sample design) can be tested adequately only after a fully operational web-based diary is developed. Therefore, we have attempted to separate studies into those which could be conducted before the web-based diary is fully developed and those that should, logically, wait until there is a viable web-based diary.

Development of Web-Based Diary

The proposed redesign calls for replacing the current paper-and-pencil diary with a web-based diary. BLS has developed a web version of the diary in a format for laptop and desktop computers, and plans to develop a web version in a format for mobile devices. BLS will test these web versions of the diary in feasibility studies during 2013 and 2014 (Jay Ryan, personal communication, February 1, 2013).

We recommend that BLS conduct usability testing as the web-based diary is developed, to help ensure that the web-based diary helps respondents provide their data accurately, quickly, and with a minimum of effort. Usability testing of the web-based diary could employ the following methods:

- An environmental scan;
- Stakeholder interviews with BLS and Census staff;
- Focus groups with participants who have used the web diaries;
- Prototype testing;
- Tests in the user experience laboratory; and
- A followup field test.

The environmental scan could include a review of the web-based applications that consumers can currently use to keep a record of their spending, such as Mint, Buxfer, and Moneydance. Westat (2011b) compiled a list of these products and their capabilities several years ago. . An environmental scan could examine the relative merits of the design of each product, and study how the CE web-based diary might incorporate the most promising design elements. The cost of this work is approximately \$15,000 if a contractor completes the work.

Stakeholder interviews could also provide important information relevant to the design of the web-based diary. BLS and Census staff members who have worked with the current paper-and-pencil diaries have observed the errors that respondents tend to make when reporting their expenditure data. Interviews with these staff members could reveal the most frequent errors that respondents have committed on paper diaries when reporting a purchased item, the price paid, or ancillary information such as whether a clothing purchase was for a child. For example, CE respondents may tend to make errors when a price is ambiguous because the price paid is affected by discount coupons, a payment plan, or a rebate offer. The web-based CE diary that we described earlier in this report can be interactive and present specific, relevant instructions concisely at the moment that the respondent needs guidance. The stakeholder interviews could help suggest the kinds of interactive instructions that are most needed in a web-based CE diary. The cost of these stakeholder interviews is \$15,000 if a contractor completes this work.

Focus groups with the people who participated in the field test of the early versions of the web diary could help reveal any deficiencies with the web diary, and suggest how those deficiencies could be resolved. The focus group moderators could ask about the difficulties that the participants

experienced with the web diary, and how they tried to overcome those difficulties. The moderator would project images of the pages of the web diary. The group participants could then discuss any problems that they encountered using each page. The moderator could make changes to the projected image of the page, working with the participants to arrive at improvements to the content and layout of each page.

These focus groups may be logistically difficult to arrange, especially if the field test participants are geographically dispersed. The focus groups will be feasible only with field test participants who live in single geographic area, and who are all available shortly after the field test. An alternative approach that may be more feasible would be to interview the participants individually using a conferencing system such as WebEx, which would permit the interviewer to speak with the participants while showing them pages from the diary. Nine interviews would cost \$20,000 if done by a contractor.

The environmental scan, stakeholder interviews and focus groups with field test participants could lead to ideas for revising the web-based diary. Developers could apply those ideas by creating prototypes for the next version of the web-based diary. Those prototypes could then be tested with about 5 test respondents. An interviewer would observe the test respondents as they entered specific expenditure data using the prototype pages. The respondent might be asked to “think aloud” to help the interviewer identify problems with the design of the prototypes. Prototype testing with 5 respondents would cost \$30,000 if a contractor completed the work.

The developers would revise the web-based diary following the prototype testing. The web-based diary would then be ready for testing in the usability laboratory. Respondents would be given a set of tasks (called “scenarios”) to carry out using the web-based diary. For example, the respondents might be asked to enter the data for a restaurant meal. The expenditure might be complex in some way; for example, non-household people may have shared the restaurant meal or the price may have been affected by a discount. The respondents would be video-recorded in the usability laboratory. Researchers would measure the respondents’ speed and accuracy in a number of different scenarios, and observe the points at which the respondents hesitated or became confused.

The goal of testing in the usability laboratory would be to study the experience of users of the web-based diary, and detect any problems that the developers may have missed. The developers would revise the web-based diary based on the findings. We suggest three rounds of iterative testing, with 5 respondents each. One set of three rounds would test versions of the web-based diary for desktop and laptop computers. We estimate costs of about \$75,000 if BLS contracts the work out.

A similar set of usability laboratory tests for the version of the web-based survey that runs on mobile devices would be conducted using 3 rounds of 5 respondents. The cost is \$75,000 if BLS contracts this work out.

After the completion of the usability testing, BLS could conduct a small field test in which CE respondents used their own technology, including their own computers and mobile devices, to report their expenditures. This field test would reveal whether the web-based diary could be successfully used for actual CE data collection. The field test would also reveal the extent to which respondents will use various technologies to enter data into the web-based diary. The design of the field test that BLS will conduct in 2014 calls for the respondents to be assigned a method for accessing the web-based survey. That is, respondents will be assigned to a desktop and laptop computer group, a tablet group, and a smart phone group. We recommend using a different design in which respondents are free to use any technology that they choose. The field test could then reveal the preferences of the respondents, as well as their success using the web-based diary. A field test with 200 CUs would cost \$200,000 if the Census Bureau did the work.

A separate development effort would be needed to create the smart phone app that CE respondents use to create mementos that remind them to report their expenditures. This app would be relatively straightforward. Respondents would first select the kind of memento they wished to create: a photo, a voice memo, a typed note, or a handwritten note. The app would then prompt the respondent to take the photo, audio-record the memo, type the note using the on-screen keyboard, or write the note onto the screen with a stylus or finger. The app would then automatically upload the memento. Developers could create a prototype of this app for testing. The initial version of the prototype need not actually upload the memento; the memento could remain stored on the smart phone. The prototype of the app could be tested in the usability laboratory, with 2 rounds of 5 subjects, at a cost of \$50,000 if BLS contracts the work out. The revised app could then be included in the field test of the web-based diary or fielded as a separate test. The results of the field test would suggest the proportion of CE respondents who chose to use the app to create mementos, the degree to which the respondents entered data in conjunction with the mementos, and the respondents' level of satisfaction with the app.

The development of an electronic diary is an essential part of the development work for the redesigned CE instruments. We therefore recommend that the evaluation of the web-based diary be given a great deal of priority. The web-based diary should be in final or near-final form when it is used in all the other evaluation studies.

The Content of the Interview and the Diary

We recommend conducting an evaluation to determine which expenditures are covered by the two recall interviews and which expenditures are covered by the diary. We have proposed that only food, alcohol, personal products, and possibly apparel be collected on the diary. However, that proposal should be evaluated to determine whether respondents can understand that only those expenditures are included and enter the appropriate data in the diary. This work would explore variations in how the interviewers describe the kinds of expenditure to enter into the diary. For example, the interviewers might explain not only the type of item but also the types of stores where the item was purchased.

This evaluation would be carried out through a series of small laboratory and field studies. In the laboratory, respondents would be presented with the task of filling out the diary, just as they would if they were enrolled in the field survey. They could then be asked to perform tasks aimed at understanding what items they would include in the diary. One task would be to list all items they believed should be included on the diary. In another task, they would examine lists of expenditure items and indicate whether they would include each item on the diary (similar to a card sort type of exercise). The dependent measure would be how closely respondent choices corresponded to the correct expenditures for the diary. We would anticipate that this work could be completed in 2 to 3 rounds of lab-based work, with a small number of respondents, about 8 per round. Another approach would be to conduct the initial work using focus groups in which the participants complete the tasks and then talk about why they included or excluded particular expenditure items on the diary. This work would cost approximately \$40,000.

We also recommend small scale field studies in which respondents keep 1-week diaries. Depending on the results of the laboratory studies, the expenditures covered in the diary or the instructions that the interviewers give the respondents for keeping the diary could be varied across respondents. In these field studies, the respondents might also be asked to complete a recall interview prior to keeping a diary, to mimic the proposed integrated design for the CE. The interviewers could debrief the respondents after the weeklong diary-keeping period, to learn how well the respondents understood and acted upon the instructions for keeping the diary.

Sample sizes in the field studies would be dependent upon the design of the field studies. Total sample size most likely will not need to exceed 100 completed diaries; costs for such a small scale field test should be about \$480,000, assuming that the work is contracted out.



The Use of Global Questions

Global questions can take two different forms. With the first kind of global questions, respondents provide a single estimate for an entire category. The second type of global question employs a filter or “gate” question. Respondents first report whether they had any expenditures in a category. They receive questions about those expenditures only if they answer affirmatively. Both types of questions should be studied.

The initial developmental research would determine the feasibility of asking respondents to provide global estimates for entire expenditure categories. The research would include both cognitive laboratory studies and small-scale field studies, building upon past work by Edgar (2011). This developmental research would suggest which expenditures are best suited for the global questions, and what strategies the respondents employ to reach a final estimation of their expenditures. For example, the research will suggest whether respondents are decomposing the expenditures into specific items or using some other strategy.

The research might also suggest how accurate the respondents’ estimates tend to be when they are asked global questions. To assess the accuracy of the respondents’ estimates of their expenditures, an interviewer could debrief the respondents and other household members about the expenditures to learn the actual levels of expenditures. The interviewer could also ask the respondents to provide receipts and other expenditure records.

We expect that this developmental research will require 3 to 4 rounds of cognitive laboratory research, each involving 6 to 8 respondents. The research would entail developing a set of global questions for the interview. Those questions could be modified after each round of testing. The cognitive lab results will suggest which OPLC categories are most suited for global questions. We expect that these categories will include, for example, large expenditures that do not require significant calculations (e.g., furniture, appliances) or regular expenses that could be estimated relatively easily (e.g., gasoline). This developmental research would cost approximately \$50,000.

Length of Reference Period

Our proposed redesigned CE interview employs reference periods of 1, 3, and 6 months for various expenditures. Several issues must be investigated with respect to this plan. One is deciding on which

expenditures can be asked for a 6, 3 and 1 month period. Second, it is important to assess how variation in the reference period during the interview might confuse respondents and lead them to telescope, or unintentionally truncate, the periods. We propose a series of cognitive interviews to study these two issues. The respondents would complete the interview using varying reference period conditions. The respondents would then be debriefed to study their reactions to the questions. The accuracy of the respondents' estimates of their expenditures would be assessed by debriefing the respondents, reviewing the respondents' receipts and other records or by asking the respondents to continually provide records of their expenditures over time. The respondents in this research could be limited to ones who were willing to carefully keep receipts and other expenditure records, so that the accuracy of their expenditure estimates could be assessed. Assuming 4 rounds of cognitive interviews with 8 respondents, the cost would be approximately \$50,000 if the work was contracted out.

Use of Annotated Grocery Receipts

We recommended allowing the respondents to provide receipts in place of making diary entries for grocery store and supermarket expenditures. Two key issues must be investigated: (1) will respondents annotate their grocery receipts in sufficient detail to permit keying and coding? (2) should data on the receipts be entered by field interviewers or by a centralized keying operation?

We recommend beginning with small tests involving about 25 respondents to determine whether respondents can annotate the receipts in sufficient detail so that the OPLC categories can be determined for the items. Respondents would keep and annotate grocery receipts for a week. Their receipts would be examined to determine whether the OPLC categories could be identified for each item purchased.

After those initial tests, more extensive tests could be built into current CED diary collection. Some CUs would be permitted to provide annotated receipts instead of making diary entries for their grocery store and supermarket expenditures. The number and the dollar amount of the expenditures could be compared when respondents are and are not allowed to provide the receipts this way.

NORC (2012) conducted related research in which a central coding operation entered data from receipts that CE respondents provided; that study, however, involved only about 150 CUs. We

recommend a larger test to help determine whether field interviewers should key the data on the receipts into a database, or whether a centralized coding operation should enter those data. There is a good reason for using a centralized coding operation: It can potentially be an efficient operation employing highly trained coders who use effective quality control measures. There is also a good reason for using field interviewers: They can contact the respondents easily to ask for more information about any item on the receipt. We recommend conducting a feasibility study that focuses on the use of a centralized coding operation. We believe that such an approach will be both more cost efficient as well as result in the highest quality data.

Moreover, with a centralized coding operation, some of the data entry might be automated. In recent years, the grocery industry has greatly consolidated. In 2009, four companies (Wal-Mart, Kroger, Costco, and Supervalu) accounted for more than 50 percent of sales in the United States, and in the largest 100 metropolitan areas, these retailers accounted for 72 percent of grocery sales (Martinez, 2007). Negotiating with these four major retailers to obtain their receipt coding scheme would in principle allow the CE program to incorporate centralized, automatic coding of receipts.

The small scale study involving about 25 respondents would cost approximately \$25,000 assuming it was contracted out. The larger test could be built into the larger field tests outlined below.

Record Use: Understanding Record Organization and Willingness to Access

A study by Sjoblom and Lee (2012) suggested that respondents were both able and willing to collect records related to the expenditures that they entered into a diary over 2 weeks. The study also suggested that it was feasible, though time consuming, for interviewers to scan the receipts.

We recommend conducting an evaluation study with the aim of identifying the best methods to motivate respondents to use expenditure records in the initial CE interview. The evaluation study would also lead to protocols for conducting the interview, both when the respondent does and does not use records. The evaluation study would identify which records respondents are likely to use, and which records respondents are unwilling or unable to access or share.

We recommend two studies to learn more about CU expenditure records. Interviews and focus groups with respondents would be beneficial to further understand how respondents store and

organize their records, the extent to which they use electronic record keeping and bill paying, and their willingness to access those records for data collection. This research could help identify the reasons that respondents might not be able or willing to retrieve particular expenditure records. Ideally, these focus groups could be conducted around the country, to achieve geographic diversity. For purposes of budgeting we have assumed 2 focus groups costing a total of \$20,000.

At the same time, we recommend that data concerning record usage be collected as part of the ongoing CE program. Interviewers could observe and record respondents' use of records for the interview. Collecting data about record usage this way would be a low cost means by which to study this important issue, for which we have little empirical data. The results would show how often respondents use records in the interview, which particular respondents tend to use records, and what sorts of records they tend to use.

We also recommend conducting an experiment to study how the advance letter sent to CUs could best motivate them to prepare for the first interview by gathering records. The experiment would compare alternative versions of the advance letter (possibly with incentives). This experiment could be conducted with a sample of CUs that are participating in the ongoing CE program. Interviewers would note the extent to which the CUs used records in the initial interview, the types of records that they used, and the specific kinds of information that they used the records to report. Sample sizes for this experiment should be large enough to provide sufficient statistical power. To compare two different advance letters, in two waves, a sample of approximately 1600 cases would be needed if 400 CUs were assigned to each condition. The costs would be marginal if this experiment were integrated with current CE data collection activities; it would be more costly if it were conducted as an experiment with its own sample.

We recommend conducting an additional experiment to compare two methods for asking respondents to use their expenditure records. In one condition, the advance letter to the CUs would ask respondents to gather their records prior to the first interview. In the second condition, the respondents would be asked to gather their records at the conclusion of the first interview. The interviewers would call the respondents a few days later to collect the data that the respondents found on their records. This experiment would identify which method is better for eliciting respondent cooperation and enhancing data quality. Once again, we believe that this work could be integrated with the current CE interview design, with marginal impact on costs.⁶

⁶ We would recommend using the current CE interview for this work, with some minor additions asked of the field staff concerning record use by the respondent. The research would require that for those CUs assigned to the post-interview revision, a second telephone call would be made to the CU to obtain updated/corrected information.



We emphasize that these developmental studies should be grounded on the philosophy outlined in Section I. The respondents should be apprised of how to best perform their tasks and the interviewers should be evaluated not just on their ability to achieve a completed interview but also on their ability to persuade respondents to keep records. A successful approach to the use of records will require that respondents and interviewers perceive their roles as a collaborative venture to achieve the highest quality data. For that reason, we recommend that the experiments and field studies incorporate the use of incentives and that field interviewers be evaluated on a full range of metrics that align with the goals of the use of records, that is, improved data quality and reduced levels of nonresponse.

Development and Testing of an Integrated Sample

One of the first evaluation activities should be an assessment of the feasibility of the integrated design. If such a design is deemed to be infeasible for any reason, then the CE program must continue to employ a two-sample design. However, the integrated design can be tested only when the web-based diary is fully developed. We recommend conducting a test of the integrated design as soon as the web-based diary is ready.

We also see merit in testing the order of the diary and the CE interview. This experiment could address the question of whether respondents should keep the diary for 1 week before the initial interview, or after the initial interview, or before the second interview, or after the second interview. However, such an experiment requires the development of different protocols for the advance letter and initial contact; in addition, it may require modifications to the CE interview. The experiment would therefore be quite expensive. For that reason, we do not recommend conducting the experiment (see last section of this chapter).

Instead, we recommend testing the integrated design in two stages: a small scale pilot test, followed by a larger scale field test. The purpose of the pilot test is to study how to encourage respondents to make diary entries without any lapses throughout the reporting period, and how to best encourage respondents to gather records after the initial interview so that they can provide detail about their responses.

In the pilot study, the respondents would complete an interview and then keep a diary over a 1 week period. Interviewers would contact CUs where some or all of the respondents appear to have

stopped making entries into their web-based diaries. The interviewers would telephone, email or text the respondents who have neither made entries on the web-based diary, nor checked the box to indicate that they had no expenditures to report, for at least 1 day. The interviewers would observe and record whether the respondents resumed making diary entries. Similarly, interviewers would contact respondents using the telephone, or email, or text message, to remind them to gather records that could improve upon their responses during the initial interview. The interviewers would observe the extent to which the respondents complied with this request.

The interviewers would debrief the respondents at the end of the pilot study. The respondents would report their reaction to the contacts from the interviewers.

As this research would be a pilot study, it should be conducted with a small number of CUs, not a large enough number to achieve statistical power. We recommend including approximately 100 households, which would cost approximately \$600,000, assuming this would be done by a contractor.

The second, larger, study would be a formal test of the recommended design. It would examine response rates, process measures (e.g., keeping up with diary entries) and the number and dollar amount of the expenditures that the respondents report, both in total and by category. This study could examine the degree to which an integrated design can yield better quality data than the existing CE program, which uses two separate samples, is currently providing. If 1,000 CUs participated in this field study, and the Census Bureau did the work, the cost would be approximately \$860,000.

Field Test: Global Questions, Diary Content and Reference Periods

We recommend conducting a field test to study the effect of employing global questions in the interview. The field test would test both kinds of global questions discussed in Section II. This field test could be conducted simultaneously with the test of the integrated design, that we recommended above.

The field work would compare the two kinds of global questions for particular sets of expenditures. Those expenditures would be identified by the earlier developmental research. The length of the reference period could be varied. Of particular interest are items that are being changed from the current design (e.g., from 3 months to 6 months, or from the diary to a 1-month reference period on



interview). The dependent measures on this field test would be the number and level of expenditures, the time needed to complete the interview, the perceived burden and perhaps respondent-interviewer behavior (e.g., expressions of frustration, reading questions as worded).

We recommend conducting this field test outside the D.C. area, preferably in 2 to 3 locations with experienced CE interviewers. With a sample size of 1,000 the cost would be around \$500,000 if the work were done by the Census Bureau.

Final Field Test

We recommend conducting a final field test to study design issues related to two issues: using multiple diary keepers in each CU and providing incentives.

Multiple versus Single CU Diary. BLS plans to conduct a study in 2014 to compare having multiple diary keepers in each CU with the current design of having one diary keeper per CU. However, that test will not incorporate many of our recommendations. For example, the diary data collection period will be 2 weeks rather than 1 week as we recommend. The electronic diary will resemble the current paper diary, and not use an interactive design as we recommend. The diary will cover all expenditures, rather than a limited set of expenditures, as we recommend.

We therefore suggest that the field test that BLS is planning should be redesigned to incorporate our suggestions for the design of the web diary, as presented in this report. The dependent measures in the field test could be response rate, and the number and dollar amount of the expenditures reported, in total and by OPLC categories.

We also recommend an experiment in the field test to study how the interviewers should contact CUs that appear to be neglecting to make entries into the web-based diary. The experiment would build upon the pilot study related to the integrated design. It could reveal whether the interviewers should contact the CUs by telephone or by email. It could also suggest the optimal way the interviewers should ask the respondents about the apparent lapse in their diaries. The experiment could also study the issue of who in the CU should be contacted. For those CUs assigned to the multiple diary condition, we believe that follow-up can only be with those CU members with whom the interviewer has had contact during the CU interview. For example, if a husband and wife jointly participate in the CE interview, the interviewer would be free to email or call either of them;

however, for the adolescent daughter or son who is living in the CU but who was not present during the CE interview, communication concerning diary entries would have to be conducted through one of the parents. Due to issues of confidentiality with respect to the diary entries, this communication would be limited to encouraging compliance.

We envision a relatively large field study in which the dependent measures included both response rates and data quality. Data quality could be measured as the number and dollar amount of reported expenditures, both in total and by OPLC category. If this field study incorporated no experimental design, sample sizes on the order of the 2014 study would be appropriate. Ideally, the field study could employ a sample size sufficiently large to include an experimental design. The experiment could provide valuable information about how the interviewers should contact CUs where some or all of the respondents appear to be failing to make entries into their web-based diaries.

Incentives. We also recommend conducting an experiment to help identify the optimum timing and dollar value of the incentives. The experimental conditions could be the following:

- Incentive with advance letter: \$0, \$2, or \$5
- Incentive at Interview: \$0, \$25, or \$50
- Incentives with Diary:
 - For the CU, if the CU has only one respondent: \$0, \$25, \$50
 - For each respondent, if the CU has multiple respondents: \$0, \$5, or \$10 with \$20 for the principal CU respondent.

We recommend that this field test be the final field test, to be conducted in year 5 of the evaluation period. By then, BLS will have completed several field tests that would suggest whether this final field test should incorporate global questions for the interview, and varying reference periods for the expenditure items on the interview. The earlier field tests would also suggest whether this final field test should allow respondents to provide receipts for their grocery and supermarket expenditures in place of making diary entries, how the web-based diary should be designed, which categories of expenditures should be covered by the diary, how respondents should be asked to use their own technology to access the web-based diary, and how respondents should be encouraged to use records. In other words, when this final field study is designed, most of the major design decisions will have been made.



Based on current Census-BLS field tests (most notably the Census Multiple Diary test), we recommend a sample size of approximately 3,000 CUs for the final field study. A sample size of 3000 completes will permit BLS to examine the impact of using multiple respondents per CU and the effect of various incentive strategies. The dependent measures can include response rates and the number and dollar amount of reported expenditures, both in total and by OPLC category.

We estimate costs for this final field study to be at approximately \$3 million, based on Census cost estimates for a number of feasibility and evaluation studies that are currently underway.

Statistical Power

Tables 4 through 6 show the power of the statistical tests for these experiments. The sample sizes for the tests proposed prior to the final test do not generally allow for extremely powerful statistical tests. When the number of completed interviews is 1,000, for measures in the form of a proportion, such as the percent that report a particular expenditure, there is between 60 percent and 70 percent power to detect a 10 percentage point difference. For example, with 1,000 completed interviews, each of 2 treatment groups would have approximately 500 respondents. Table 4 shows that for proportions of approximately 50 percent, the level of power is about 60 percent to detect a difference of 10 percentage points.

The sample sizes double when comparing measures to the ongoing Census production sample. This is because the entire field test sample can be used for estimation. For example, when comparing the integrated design to the production sample on the pre-notification procedures, 1,000 completed interviews would provide 90 percent power to detect a 10 percentage point difference. But the power to detect a 4 percentage point difference would still not be large.

Of course, the power for the final field test is much larger because it involves 3,000 completed interviews. The power when comparing mean expenditures varies widely depending on the particular category (Tables 5 and 6). Samples that provided approximately 1,000 completed interviews would yield sufficient power to detect differences of about 20 or 30 percent.



Table 4. Power for difference in proportions by sample size, level of rate and percentage point difference between the groups*

Completes	p=50%		p=70%		p=80%	
	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%

*Power calculated for a two-sided significance test at $p < .05$. assumes a design effect of 2; +percentage point difference.

Table 5. Power for difference of means for selected expenditures, completed diaries per group and effect size*

Expenditure type	N = 500			N=1,000			N=2,000		
	10%+	20%	30%	10%	20%	30%	10%	20%	30%
Food at home	14%	43%	76%	24%	71%	96%	43%	94%	99%
Cereals and cereal products	10%	28%	56%	17%	51%	84%	29%	80%	98%
Eggs	9%	22%	43%	13%	39%	72%	22%	67%	94%
Other food at home	12%	33%	63%	19%	58%	90%	34%	87%	99%
Food away from home	13%	38%	70%	21%	65%	94%	38%	91%	92%
Alcoholic beverages	5%	10%	17%	7%	15%	29%	10%	26%	51%

*Power calculated for a two-sided significance test at $p < .05$. Based on 2011 expenditure data.

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

Table 6. Power for difference of means for selected expenditures, completed interviews per group 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 $p < .05$. Based on 2009 expenditure data.

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

Summary of Schedule and Costs of Development and Testing Plan

The schedule for the evaluation is summarized in Figure 7. The evaluation will require a total of 5 years (60 months). We recommend conducting the tests on several tracks. The first track pertains to the tests of the respondents' use of records. These tests will require 36 months in total, counting



all of the work outlined above. As these studies are conducted, the findings about how respondents can best be encouraged to use records can be applied to the subsequent evaluation studies.

The second track relates to the studies to identify the expenditures that should be covered by the interview and by the diary, the use of global questions and variations in the reference periods. The preliminary studies should be completed by around month 18, with the field test starting in month 19 and lasting through month 30.

The third track relates to the development of the integrated design. The initial work requires the development of the web-based diary, which should last through month 25. The testing of the integrated design would start in month 26 and last through month 40, when the large field test would end.

Figure 7. Timeline for redesign tasks

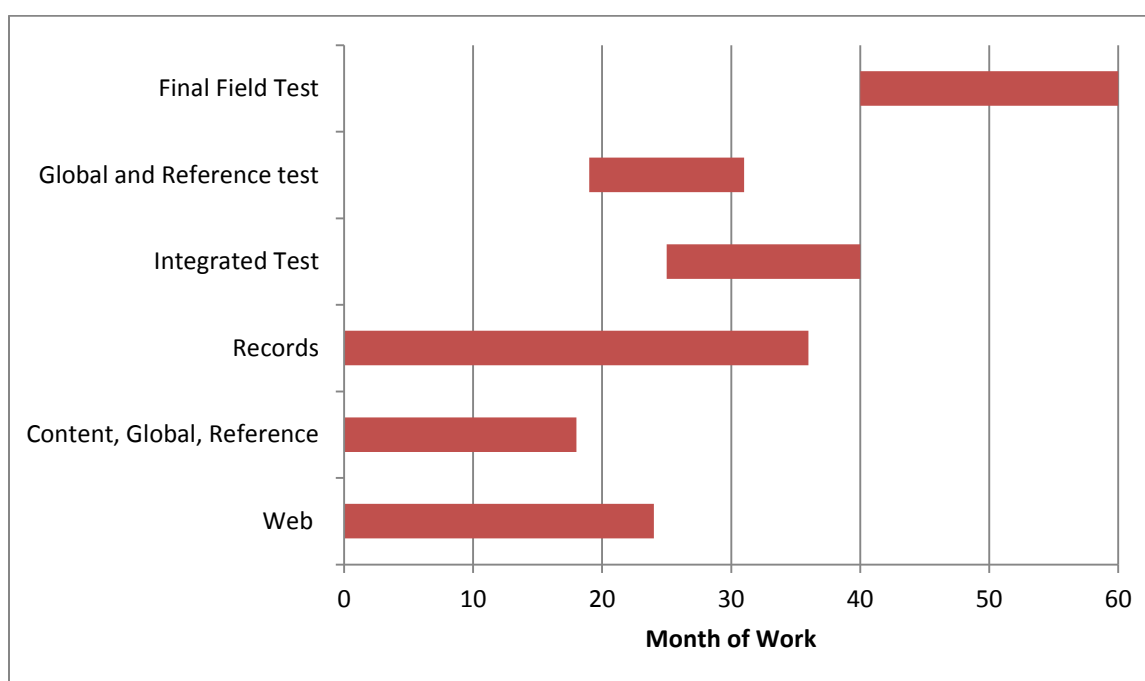


Table 7 summarizes the estimated costs of the design. These vary, depending on the extent that BLS or a contractor conducts the work. If BLS conducts all of the preliminary lab-based work in-house, the cost is approximately \$5.5 million. If all of this work is contracted out, the cost is approximately \$6 million.



Table 7. Summary of expenses for R&D work (\$1,000)

R&D task	Census	Other work		Total	
		If contracted out	If done by BLS	If contracted out	If done by BLS
Technology development	\$200	\$280	0	\$480	\$200
Content	—	\$520	\$480	\$520	\$480
Records	—	\$45	0	\$45	\$0
Integrated design	\$860	\$600	\$600	\$1460	\$1460
Global questions and reference period	\$500	\$100	0	\$600	\$500
Final test	\$2,900	0	0	\$2900	\$2900
Total	\$4,460	\$1,545	\$1,080	\$6005	\$5540

Scope of the Research and Development Program

The CE has not been redesigned for over 30 years. This may be one reason why many of the new designs suggested by the NAS panel and others have included very significant changes to the current methodology (Dillman and House, 2013; Westat, 2011; Mathiowetz et al, 2011). These designs have all tried to address two basic issues. One is to reduce the overall burden on CE respondents. Many have pointed to burden as one of the root causes of measurement error and response rate issues associated with the survey (e.g., Lessler, 1987). To reduce the burden, the basic structure of the CE has to be reconsidered. The second issue is related to the evolution of computer assisted interviewing (CAI). The CE currently uses CAPI, but the landscape for data collection is evolving very rapidly. Many respondents are able to access their own computers or mobile devices, which should make it easier to collect data on expenses. This should improve data quality.

The research and development program that we propose targets the main changes to the current design, including integrating the interview and diary components, the adoption of a web-based diary, changing the content of the two surveys and changing the number of interviews per household. To fit within the financial constraints, the proposed research and development program does not allow for as much exploratory research as is optimal. For a number of the key decisions, the plan assumes it is possible to narrow down the new methodology to a single design without a significant field test. At several points, the development could benefit from testing alternative methods against one another before narrowing to a final design. Examples include:

1. Testing several different versions of the web diary. The assumption in the current plan is that a single version will emerge from the initial development and testing. This



may be premature. It would be better to consider the possibility that several different versions may seem promising and carried forward to more testing.

2. Testing several different versions of the content of the interview and diary surveys. The plan assumes that an optimal single design will emerge from the preliminary testing.
3. Alternative spacing of the two interviews. One key feature of our design is the recommendation to incorporate two rounds of CE interviews, 12 months apart. The 12-month lag between interviews – albeit a reduction in total burden placed on the CE unit – will require the development of field methods to maintain a high response rate for the second interview. Alternative spacing, such as 3 months or 6 months, also has implications for the recall of expenditures and the analytic utility of the data.
4. Integrating the sample design. For reasons stated in chapter 2, we have recommended integrating the interview and diary samples. The testing plan has a final test for this plan. However the plan does not allow for much time (or resources) to experiment with different ways to integrate the sample. For example, there was considerable discussion about whether the diary should be administered prior to conducting the interview. In our judgment, doing the interview first will provide respondents with a better idea of what is needed when completing the diary. This order should also be less redundant than the other way around. However, this is not based on empirical data and could benefit from more work at the development stages.

This list is purely illustrative of the types of studies for which the R&D effort could benefit. It is not meant to be exhaustive or in any priority order. Each of these activities would add an additional round of field testing to narrow down the alternatives before conducting a larger field test that compares the results against the current design.

A second issue related to cost is the sample sizes proposed for the larger field efforts. In particular, it would be prudent to significantly increase the size of the tests. As shown above, the power to detect differences in mean expenditures is not particularly high.

We fully understand that R&D has to be done within an environment of fixed funding and competing demands. Nonetheless, spending \$5 to \$6 million on the redesign of a major survey like the CE, which has data collection costs of approximately \$30 million per year, strikes us as low. We would urge allocating more resources if at all possible.

A major constraint in creating the revised design was the requirement that it be cost neutral. As we considered various design elements, we kept this requirement firmly in mind, and endeavored to avoid designs that would clearly be too costly. However, we did not continually develop cost estimates as we developed our recommendations for redesigning the CE program. Rather, we carried the design work to a nearly final state before we began to estimate the cost of the revised design.

Another constraint was to develop the cost estimate so that it would represent the Census Bureau's cost for conducting the revised design. BLS informed us at the start of this project that a decision had been made to continue working with the Census on the CE, throughout the development and implementation phases of the redesign and beyond. Thus, we were faced with the task of developing cost estimates for another organization to perform the work, with only high-level summary information about their costs from the current design. The Census Bureau's data collection operations and cost structure differ from other national survey organizations in the U.S. in a number of important ways. In consultation with BLS, we decided not to stray far from the summary CE cost data that Census provided, but to supplement it with our own understanding of Census operations on projects like the CE and our own cost experience on similar projects where necessary.

The timeline for design and implementation of the revised CE program is about 10 years. The cost estimate is intended to apply to a typical year beyond that, after the implementation was complete. Creating cost estimates was quite challenging because the CE is a complex set of activities and technologies, and 10 years is beyond the horizon for estimating survey program costs on most projects.

Objectives

The cost exercise had three objectives:

1. To estimate the annual cost of the revised design in its steady state, after implementation;

2. To create assumptions based on recent cost data provided by the Census Bureau; and
3. To specify the assumptions used for each cost element.

The Bureau of Labor Statistics provided summary cost data from the Census Bureau for conducting the CE in 2011. Total cost was about \$30,000,000. This cost included both the CED and CEI components. It did not include any program costs at BLS. These figures are not actual costs but represent average cost estimates for a typical year of CE operations, rounded to the nearest \$100,000 level.

The Census costs were broken out into four categories:

- | | | |
|----|----------------------------|----------------|
| 5. | Total field activity tasks | \$6.4 million |
| 6. | Interview only | \$10.7 million |
| 7. | Diary only | \$3.5 million |
| 8. | Overhead | \$9.2 million |

Interview costs were also specified by wave and interview status (complete or non-complete) for each of the five waves in the current design, and mileage was given for the “interview only” category. No additional cost detail was provided.

Assumptions and Limitations

The first assumption we made was to provide costs in 2011 dollars. BLS and Census can choose to apply an inflation factor to convert to equivalent costs in later years.

Another assumption was that, wherever our own cost experience on similar projects differed from the data provided by the Census Bureau (for example, on cost relationships between waves of a panel study), we accepted the Census cost data as the basis for estimating the redesigned CE costs. This assumption is a limitation of the redesign estimate; its quality is dependent on the Census 2011 data, which have very little detail and no supporting documentation.

A standard approach to developing cost estimates is to identify fixed and variable costs. Clearly interviewing costs vary with sample size, and the cost of salaried labor for headquarters staff is largely fixed, regardless of sample size (over a relatively broad range of sample sizes). However, this

division of cost categories is not clearly delineated in the CE cost data provided by the Census Bureau. Therefore, we inferred fixed costs from the brief labels and descriptions that accompanied the cost data, and focused attention primarily on the costs we knew were mainly variable.

The Census provided no information about the structure of their fixed direct and indirect costs. Therefore, we made the simplifying assumption that the indirect costs for the redesign would be similar to the 2011 overhead. In 2011, Census overhead was \$9.2 million, about 45 percent of direct costs.

The fixed part of direct costs was also not defined, but we assumed they were included in the category “total field activity tasks.” The Census Bureau gave three examples of activities included in this category: re-interviews; senior field representative conference; and address research. These examples suggest that activities included in this category may not be specifically related to the volume of data collection (in that they can be scaled to fit a pre-determined need, somewhat independent of sample size), and therefore might be considered fixed. However, we have no way to estimate what mix of activities might be included in such costs after implementation. Total field activity costs were \$6.4 million in 2011, 45 percent of the interviewing and diary costs, so we assume they would be 45 percent after implementation of the redesign.

Extrapolation of Essentially Variable Costs from Census Cost Information and Westat Experience

The variable costs in 2011 were divided between “interview only” and diary. Our approach for estimating these was to:

- Identify cost elements that can be mapped between the current design and our recommended redesign, or that are only in the redesign;
- Plug in design features of the redesign (e.g., sample size, number of waves); and
- Apply common cost metrics.

We based the cost metrics on our own experience with similar surveys, but made some adjustments for what we know about the Census field staffing configuration and the level of Census field effort required for the current CE.

Table 8 compares the cost elements in the current design with the redesign. The redesign has 28,000 CE interviews per year, 7,000 fewer than the current design because there is no separate bounding interview. The integrated redesign has no additional sample of CUs for the diary component, whereas the current design has 7,000 additional CUs for the diary.

Table 8. Cost elements shared between current design and proposed design

	Current design	Proposed design
CUs interviewed	7,000	14,000
Interviews completed	35,000	28,000
Bounding interview	7,000	—
Data interview	28,000	28,000
Diary		
CUs	7,000	No additional CUs in integrated design
Diary weeks	14,000	14,000

The variable costs related to interviewing in the integrated redesign have 7 basic components:

- Interviewer labor;
- Mileage costs;
- Senior interviewer labor;
- Survey supervisor labor;
- Respondent tablets;
- Respondent incentives; and
- Handling and transcribing grocery receipts.

Interviewer labor is the most costly of these components. To estimate interviewer labor for the redesign, we made use of data on contact attempts BLS provided from the current design, drawn from Census paradata. Table 9 shows mean attempts by wave for the current design. A total of 23.7 attempts are made over 5 waves to complete CE interviews with the average CU. This includes both in person and telephone attempts. Almost all of the first wave interviews have at least one in-person attempt, but many interviews in later waves are conducted completely by phone, with no in-person visit. The number of contact attempts by mode was not provided to us.

Table 9. Field representative labor: Contact attempts per consumer unit

Wave	Current design (without diary)	Proposed redesign (with diary)
1	5.4	10.0
2	5.0	6.0
3	4.6	N/A
4	4.4	N/A
5	4.2	N/A
Total	23.7	16.0

The redesign has only two waves of interviewing per CU, 3 less than the current design, but in the redesign, the data collection during the two waves require more interviewer effort than any two waves of the current design. The additional effort in the redesign includes the following:

- **First Interview.** The interviewers conduct the recall interview and place the diary. Because there is higher burden for the respondent, more interviewer attempts are required to gain cooperation.
- **First Interview.** The interviewers track the respondents' diary entries over the week, and call when no expenditures are being entered.
- **First Interview.** The interviewers telephone diary-compliant respondents to close out the diary.
- **First Interview.** Interviewers telephone CUs to collect any corrections or additional information that the respondents found on records for the recall interview (to enhance the data collected in the interview itself).
- **Second Interview.** Interviewers conduct the recall interview in-person.
- **Second Interview.** Interviewers follow up on any records or expenses obtained after the interview (to enhance the data collected in the interview itself).

Importantly, a large majority of the Census cases across all five waves in the current design are completed with at least some data collected by telephone, although the amount of data collected by telephone was not reported by the Census Bureau. Telephone data collection is much less expensive than in-person data collection. It is not possible to compare “apples to apples” in contact attempts by mode between the current design and the redesign. However, we also note that, although in the redesign both waves require an in-person interview (unlike the current design), the number of contact attempts in the redesign also reflects a considerable amount of telephone work directly with the respondent. We have made the simplifying assumption that the mix of telephone and in-person work in the redesign's two waves is comparable overall to the mix in the current design across its five waves.

The additional effort required by the redesign in its two waves (compared to the effort required by the current design in its first two waves) is offset somewhat by the addition of respondent incentives. We estimate the first interview will require an average of 10 contacts (almost twice as many as the current design), and the second will require 6 contacts on average, one more than the current design. Without incentives, the number of contact attempts in the redesign would be greater.

We assume each contact attempt will require 2.5 hours of interviewer time. This estimate does not imply literally that the interviewer will spend 2.5 hours on each CU contact attempt. Rather, it is the result of dividing total interviewer hours on the survey by the total number of contact attempts made, across all interviewers and all cases (not just completed interviews, but all sampled cases). This produces a convenient metric that can be useful in estimating and monitoring costs.

For the redesign, we will complete 28,000 interviews, with half being a first contact (10 contacts per complete) and half being a second contact (6 contacts per complete). At 2.5 interviewer hours per contact attempt, we have 560,000 hours of interviewer time per year. We assume \$19 per interviewer hour, so the total labor cost for interviewers is \$10.6 million.

In addition to labor costs, interviewers have expenses while working in the field. By far the biggest component of field expenses is mileage; it typically accounts for 90 percent or more of field expenses. From the cost data provided by Census, it appears that the total number of interviewer miles driven divided by the number of CE completed interviews in 2011 was about 75. This is consistent with our experience with similar study designs. In 2011, the government mileage rate was 51 cents per mile. Applying these assumptions to the redesign, we have 28,000 completed interviews times 75 miles per interview at \$.51 per mile, or about \$1.1 million in mileage cost.

Senior interviewers can be viewed as the first level of interviewer supervision in the field. Senior interviewers coach new interviewers, help them resolve problems, and have specialized refusal conversion skills. We assume a ratio of 1:20 for senior interviewers to interviewers.

We estimated the number of interviewers using the following calculation:

- 560,000 interviewer hours per year;
- Interviewers work an average of 10 hours per week on the CE, for 48 weeks each year, so each interviewer averages 480 CE hours per year; and

- 560,000 total interviewer hours divided by 480 hours for one interviewer = 1167 interviewers.

Applying the ratio of 1:20 results in an estimated 58 senior interviewers. Assuming the senior interviewers work 20 hours per week on the CE for 48 weeks each year, at an average rate of \$25, we arrive at an estimate of \$1.4 million for senior interviewer labor.

Survey supervisors are the next layer of field management. These supervisors may have responsibility for 100 interviewers or more. Senior interviewers report to the survey supervisors. We assume a ratio of 1:4 for survey supervisors to senior interviewers, so with 58 senior interviewers we would have 15 survey supervisors. If we assume the survey supervisors work 20 hours per week and 48 weeks per year, at an average rate of \$30 per hour, we arrive at an estimate of \$.4 million for survey supervisor labor.

The redesign anticipates that each respondent in a CU will use a tablet, laptop, or other computing device for entering diary data. (This is a unique feature of the redesign, not included in the current design.) The majority of respondents will choose to use their own technology, but for those who do not, or who have no computing device, we assume the interviewer will provide a tablet. We assume 40 percent of the 14,000 CUs will house 1.5 respondents who require a survey-supplied tablet, so 8,400 respondents would need tablets supplied by the project. They will retain the tablets for less than a month, so during the year there would be a continuous flow of tablets going out to CUs and coming in from CUs. We assume that peak usage in any given month would be about one-sixth of the 8,400, so no more than 1,400 tablets would be deployed or in transit at any point during the year.

We assume a monthly charge of \$50 for these tablets. This includes amortization of their purchase cost over their expected life (perhaps \$15 per month), and software license fees, data charges, and maintenance costs averaging perhaps \$35 per month. The total estimated cost per year is 1,400 tablets at \$50 each per month for 12 months, or \$.8 million.

Respondent incentives are another unique feature of the redesign, not used in the current CE. For costing purposes we assume a \$2 incentive in advance, \$25 per completed interview, and \$10 per completed diary with 2 diaries per CU. With 14,000 initial contacts ($\$2 + \$25 + 20 = \$47$) and 14,000 second interviews ($\$2 + \$25 = \$27$) the total is around \$75 per CU ($\$47 + \$27 = \$74$). For 14,000 CUs this would total about \$1.1 million.⁷ We have factored the benefits of these incentives

⁷ Note that with 12 months between interviews, and once the rotation schedule is fully balanced, there would be 28,000 CU's interviewed in any particular annual period.. Half the CU's will be experiencing a first contact and half a second contact..

into our assumptions about interviewer level of effort to achieve response rates comparable to the current CE. Without incentives, we would need to assume that more contact attempts would be needed per completed interview. With some experimentation it might be possible to identify a cost-neutral point at which the tradeoff between incentive cost and interviewer labor cost would be zero, but the literature on incentives in surveys suggests that incentives are cost-effective. That is, incentives can increase response rates more cheaply than additional interviewer labor, up to some threshold.

Table 10 summarizes the estimated Census interviewing costs in the redesign. The total cost is estimated at \$15.3 million.

Table 10. Estimated Census interviewing costs in redesign

Component	Estimate
Field representation labor	\$10.6 million
Mileage costs	\$1.1 million
Senior field representative labor	\$1.4 million
Field supervisor labor	\$.4 million
Respondent tablets	\$.8 million
Respondent incentives	\$1.1 million
Total interview costs	\$15.3 million

Table 11 shows the total estimated annual costs for the redesign in a steady state at Census.

Table 11. Total costs for redesign in steady state at Census

Interviewing costs	\$15.3 million
Other field activities @ 45% of interviewing costs	6.9 million
Total direct costs	22.2 million
Overhead @ 45% of direct costs	10.1 million
Total costs	32.2 million

Benchmarking and Sensitivity

At the suggestion of BLS, we checked our assumptions against the current CE interview. We used the same cost metrics from the redesign, but applied them to the current design, with 7,000 CUs, 5 waves, 24 contacts per CU per year, and no incentives or respondent tablets. The result was an estimate of \$9.3 million, very close to the cost data the Census provided for the CE interview (\$9.4 million).

By far the single most important cost driver in our estimate is the number of contacts per CU. Changing the assumption for number of contacts per CU per year from 16 to 17 would increase the total cost by about \$2.0 million. Contact data is notoriously imperfect, but we assume the same error properties from year to year, and from one design to another in the Census contact record data.

We conclude that the redesign costs are estimated to be approximately the same as the current design costs. Our estimate for the redesign costs is very sensitive to the degree of field effort required to achieve the desired number of completed interviews and diaries and the response rate targets.

As part of these cost activities, we have produced a set of metrics and assumptions that can be modified as the redesign effort unfolds over the next several years, and as BLS acquires more cost experience with the redesigned CE program.

References

- Baumgartner, R., and Rathburn, P. (1997). "Prepaid Monetary Incentives and Mail Survey Response Rates." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Norfolk, VA, 1997.
- Beatty, P. (2010). "Considerations Regarding the Use of Global Survey Questions." Presentation at the Consumer Expenditures Survey Methods Workshop, December 8-9, 2010, Suitland, MD.
- Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., and Yamamoto, K. (1992). "An Experiment in Monetary Incentives." American Statistical Association, Proceedings of Survey Research Methods Section, Alexandria, VA: 393-398.
- Bureau of Labor Statistics. (2009). "Survey Source of Data for Consumer Expenditures Survey Integrated Tables, 209. Available at http://www.bls.gov/cex/ce_2009source.pdf.
- Bureau of Labor Statistics. (2010). "Recall Period In the Consumer Expenditure Surveys Program." Presentation at the CE Methods Workshop, December, 2010, Washington, DC.
- Cantor, D. (2010) "Discussion of Plans for Designing the Recall Period for the Consumer Expenditure Interview Survey." 2010 Consumer Expenditure Survey Methods Workshop, U.S. Bureau of Census, December 8 and 9.
- Casey, W. (2010). "CPI Requirements of CE." Washington, DC: Office of Survey Methods Research, Bureau of Labor Statistics, U.S. Department of Labor.
- Cho, M., Eltinge, J., and Steinberg, B. (2004). "Temporal Patters of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview and Other Panel Surveys." Paper presented at the Annual Conference, American Association for Public Opinion Research.
- Chopova, B., Edgar, J., Gonzalez, J., King, S., McGrath, D., and Tan, L. (2008). "Assessing Nonresponse bias in the CE Interview Survey: A summary of four studies." Washington, DC: Office of Survey Methods Research, Bureau of Labor Statistics, U.S. Department of Labor.
- Church, A. (1993). "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly*, 57:62-79
- Creech, B., Smith, M., Davis, J., Fricker, S., Tan, L., To, N., and Gonzalez, J. (2012). *Measurement Issues Study Final Report*. Bureau of Labor Statistics.
http://www.bls.gov/cex/ce_statpapers.htm.
- Davis, J., Stinson, L., and To, N. (2002). *What does it really mean to be user-friendly when designing an expenditure diary?* Presented to the American Association for Public Opinion Research, St. Pete Beach, Florida.

- Dillman, D., and House, C. (2013). (eds.) *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. Washington, DC: National Academies Press.
- Dippo, C., Coleman, J., and Jacobs, C. (1977). "Evaluation of the 1972-73 Consumer Expenditure Survey." Proceedings of the Section on Social Statistics, Alexandria, VA: American Statistical Association.
- Duggan, M., and Rainie, L. (2012). *Cell phone activities 2012*. Washington, DC: Pew Research Center's Internet & American Life Project.
- Edgar, J. (2005). "CE Data User's Survey." Office of Survey Methods Research, Bureau of Labor Statistics.
- Edgar, J. (2010). "Respondent Record Use in the U.S. Consumer Expenditure Interview Survey." Paper presented at the Annual Conference, American Association for Public Opinion Research, May 13-16, 2010, Chicago, IL.
- Edgar, J. (2011). "CE Methods Workshop Report." Washington, DC: Office of Survey Methods Research, Bureau of Labor Statistics, U.S. Department of Labor.
- Edgar, J., Davis, J., Spell, S., Verlander, R., and Wetzel, G. (2006). Individual Diary Feasibility Test Final Report. Bureau of Labor Statistics unpublished internal report.
- Fu, S.J., Wang, R., and Qiu, Y. (2002). "Daily Activity and Internet Use in Dual-Earner Families: A Weekly Time Diary Approach." *IT & Society*, Vol. 1: 37-43.
- Gelsen, E., Richards, A., and Strohm, C. (2011). "U.S. Consumer Expenditure Records Study." Paper prepared for the U.S. Census Bureau, RTI International, Research Triangle Park, NC.
- Goldenberg, K.L., McGrath, D., and Tan, L. (2009). "The Effects of Incentives on the Consumer Expenditure Interview Survey." Paper presented at the Annual Conference, American Association for Public Opinion Research, Ft. Lauderdale, FL.
- Grootaert, C. (1986). "The Use of Multiple Diaries in a Household Expenditure Survey in Hong Kong." *Journal of the American Statistical Association*, Vol. 81 (396): 938-944.
- Groves, R., Singer, E., and Corning, A. (2000). "Leverage-Saliency Theory of Survey Participation." *Public Opinion Quarterly*, 54: 346-361.
- Henderson, S., Passero, B., Rogers, J., Ryan, J., and Safir, A. (2011). "Consumer Expenditure Survey (CE) Data Requirements."
- Hicks, W., Connor, S., Ward, P., DiGaetano, R., Goksel, H., Kumar, V., and Lefkowitz, D. (2011). "An Experiment Evaluating Different Respondent Monetary Gifts." Paper presented at the Annual Conference, American Association for Public Opinion Research, Phoenix, AZ.
- Horsfield, G. (2010). Family Spending: A report on the 2009 Living Costs and Food Survey. Office for National Statistics report.

- Iida, M. et al. (2012). Using diary methods in psychological research. In: Cooper, H. (Ed.) *APA handbook of research methods in psychology: Vol. 1. Foundations, planning, measures, and psychometrics*. Washington, DC: American Psychological Association.
- Kreuter, F., McCulloch, S., Presser, S., and Tourangeau, R. (2011). “The Effects of Asking Filter Questions in Interleaved versus Grouped Format.” *Sociological Methods and Research*, 40(1): 88-104.
- Krishnamurthy, P. (2008). Diary. In: Lavrakas, P. *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.
- Lessler, J. (1987) *BLS Conference Report*. Report to the Bureau of Labor Statistics, March 12, 1987. Last accessed June 14, 2013: <http://www.bls.gov/cex/cesrvymethslessler1.pdf>.
- Martinez, S. (2007). “The U.S. Food Marketing System: Recent Developments 1997-2006.” Economic Research Report Number 42. Washington, DC: Economic Research Service.
- Mathematica Policy Research. (2011). “Supporting Justification for OMB Clearance of the National Household Food Acquisition and Purchase Survey.” Contract Number AG-3K06-D-09-0212/GS-10F-0050L. December 5, 2011.
- McGrath, D., McDermott, N., Chopova, B., and Gomes, C. (2007). “The Effects of Incentives on the Consumer Expenditure Diary Survey.” Washington, DC: Bureau of Labor Statistics, Internal report.
- McGrath. (2005). “Comparison of Data Obtained by Telephone versus Personal Visit Response in the U.S. Consumer Expenditures Survey.” Alexandria, VA: Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 3368-3375.
- O’Hare, B. (2008). “Incentives in the Private Sector.” Paper presented to the Council of Professional Associations on Federal Statistics, March 10, 2008.
- Office of National Statistics. (2011). “Living Costs and Food Survey User Guide.” ONS Social Surveys.
- Rainie, L. (2012). *25% of American adults own tablet computers*. Washington, DC: Pew Research Center’s Internet & American Life Project.
- RTI. (2002). “2001 National Household Survey on Drug-Abuse Incentive Experiment. Prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-98-9008.
- Safir, A. (2011). “Survey Design and Data Collection.” Presentation to the CNSTAT Panel on the Consumer Expenditure Survey, February 8, 2011.
- Safir, A., and Goldenberg, K. (2008). “Mode Effects in a Survey of Consumer Expenditures.” Alexandria, VA: Proceedings of the American Statistical Association, Section on Survey Research Methods, 4436-4443.

- Silberstein, A., and Scott, S. (1991). "Expenditure Diary Surveys and Their Associated Errors" Chapter 16 in P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds) *Measurement Errors in Surveys*. New York: Wiley and Sons.
- Singer, E. (2002). "The Use of Incentives to Reduce Nonresponse in Household Surveys" in R. Groves, D. Dillman, J. Eltinge, and R. Little (eds) *Survey Nonresponse*. New York: Wiley.
- 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.
- Sjoblom and Lee. (2012). "Pilot Test Report: Records Information and Feasibility of Use Study." Chicago: NORC at the University of Chicago, July 9, 2012.
- Slesnick, D.T. (2001). *Consumption and Social Welfare. Living Standards and Their Distribution in the United States*. Cambridge: Cambridge University Press.
- Smith, A. (2012). *17% of cell phone owners do most of their online browsing on their phone, rather than a computer or other device*. Washington, DC: Pew Research Center's Internet & American Life Project.
- Stone, A.A. et al. (2002). Patient non-compliance with paper diaries. *BMJ*, 324, 1193-1194.
- Tremblay, J., Lynch, J., and Dubreuil, G. (2011). "Pilot Survey Results from the Canadian Survey of Household Spending Redesign." Paper Presented at the Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.
- Tucker, C. (1992). "The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey." *Journal of Official Statistics*, 8(1): 41-61.
- U.S. Department of Transportation Federal Highway Administration. (2011). "2009 National Household Travel Survey Users Guide." Washington, DC.
- Verbrugge, L.M. (1980). "Health Diaries." *Medical Care*, 18:73-95.
- Westat. (2011a). "Redesign Options for the Consumer Expenditure Survey." Final report prepared for the National Academy of Sciences, NAS Solicitation Number DBASSE-004950-0001-031411. Washington, DC.
- Westat (2011b). *Data capture technologies and financial software for collecting consumer expenditure data*. (Report submitted to the Bureau of Labor Statistics). Rockville, MD: Westat
- Zickuhr, K., and Smith, A. (2012). *Digital differences*. Washington, DC: Pew Research Center's Internet & American Life Project.