Redesign Options for the Consumer Expenditure Survey

Nancy Mathiowetz
Kristen Olson
Courtney Kennedy

Nancy Mathiowetz is Professor, University of Wisconsin-Milwaukee. Kristen Olson is Assistant Professor, University of Nebraska-Lincoln, and Courtney Kennedy is Vice President, Abt/SRBI.

The opinions expressed in this document are the responsibility of the authors.

I.   Introduction

The Consumer Expenditure Survey (CE) is not a single survey but rather an estimation program based on two distinct data collection efforts, the Quarterly Consumer Expenditures Interview (CEI) and the Consumer Expenditure Diary (CED), each with its own separate sample. In both the CEI and the CED one household member often serves as the respondent/recorder of information. CEI sampled household units are interviewed once per quarter for five consecutive quarters; the respondent is asked to report retrospectively on consumer unit expenditures for the past three months (for interviews two through five). The initial interview collects demographic and family characteristics, an inventory of major durable goods, and expenditures for the past month, all information to be used as bounding data for future interviews. The CED is a product-oriented, prospectively-placed diary with two one-week daily expense records serving as the primary means of capturing detailed descriptions of expenses for all members of the household. Estimation of total consumptions is based on an integrated approach drawing upon data from each of the two components (see Appendix C, Survey Source of Data for Integrated Tables, 2007; BLS, 2007 for detailed information as to the data source for specific consumer expenditures).[1]

The Bureau of Labor Statistics (BLS) has undertaken a major redesign of the CE surveys. As part of this multi-year process, BLS has contracted with the National Academy of Sciences' Committee on National Statistics (CNSTAT) to review and recommend design options for the CE. The recommended redesign is intended to address (a) underreporting of expenditures in the household surveys; (b) changes in the survey environment – both the environment of consumer purchasing and the data collection environment; and (c) the need for increased flexibility in approaches to data collection.

---

[1] We note that there are redundancies in the interview and diary data collection instruments with respect to consumer expenditure items as well as for demographic information resulting in additional unnecessary burden, when viewed from the perspective of total burden across all respondents.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

A. Understanding the Problem

Measurement Error

Comparisons between estimates based on CE and the Personal Consumption Expenditures (PCE) data from the National Accounts suggest deterioration in the quality of the CE data, especially over the past decade (e.g., Slesnick 2001).[2] The source of this increasing disparity between the two data sources is often attributed to measurement error. For example, persons interviewed by telephone for the CEI have higher family income, but have similar levels of expenditures as those interviewed in person (Safir and Goldenberg 2008; McGrath 2005) suggesting differential underreporting as a function of mode. In the CED, empirical data indicate declines in reports of expenditures across the reporting period (Silberstein and Scott 1991).

Another factor that may be contributing to the disparity is nonresponse error. The period of increasing discrepancies between the CE and the PCE estimates has also been a period of fluctuating response rates (Goldenberg, McGrath and Tan 2009; Safir 2011). More important than the nonresponse rate is the strength of associations between response propensity and the expenditures being measured. For example, Abraham, Presser, and Helms (2009) suggest that nonresponse bias has led to inflated estimates of volunteerism in the U.S. because survey respondents report much more volunteering than nonrespondents. It is plausible that a similar dynamic may be undermining CE expenditure reports. Although an assessment of nonresponse bias in the CE interview concluded that there was little evidence of significant nonresponse bias, that study did find differential participation as a function of race, income and dwelling unit characteristics as well as evidence of nonresponse bias for particular expenditure categories but not for total expenditures (Chopova et al. 2008). However, in none of the assessments were actual expenditures in relevant categories from Wave 1 nonrespondents obtained, and as such, no direct evaluation of nonresponse bias was conducted.

---

[2] 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.

Because of this important limitation to these nonresponse bias studies, we hypothesize that some of the differences between the CE and the PCE may be due to differential participation and/or diary compliance. Designs that are most effective for future consideration, therefore, may be those that simultaneously attempt to reduce both measurement error and nonresponse error.[3]

Both the interview and diary data collection efforts are burdensome; the interview with respect to the cognitive burden of retrospective recall for three month reference periods,[4] the diary with respect to the sheer recording burden for a diary that may ill-fit the consumption patterns of most individuals in the 21st Century. To the extent that technology (alone or in conjunction with record keeping) can be leveraged so as to maximize the capture of consumer expenditures in real time or via a diary method – while also reducing burden in the sense of extremely long interviews or onerous recording of detailed expenditures – one would anticipate an improvement in data quality.

Technology may also facilitate the collection of information from multiple members of the household. Empirical evidence suggests that reducing reliance on a single consumer unit reporter for the diary may improve reports of expenditures. Grootaert (1986) found that expenditures, especially for personal items such as clothing, footwear, and services, were significantly higher in households assigned to a multiple diary treatment. Similarly, a field study conducted by BLS in 2006 (Edgar et al. 2006) found that the use of multiple diaries per consumer unit resulted in an increase number of expenditure items reported as well as an increased dollar value of the expenditures.[5]

---

[3] Olson (2011) notes the following: As another error sensitive indicator, we examined various proxies for measurement error. These measurement error indicators were more strongly associated with participation propensities than expenditure categories. Thus, we recommend that BLS target resources to improve the measurement error properties of those with low participation propensities. Since many of these strategies have beneficial influences on both nonresponse and measurement error….targeting resources based on measurement error may also improve participation propensities.

[4] The effect of the length of recall period on errors of omission is well documented in survey methodology literature. See, for example, the summary provided in Bound, Brown, and Mathiowetz (2001).

[5] 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

Changes in the Survey Environment

Survey designs are rapidly moving to integrate technology for purposes of improved data quality and reduced respondent burden. We see examples of smartphone data collection across a diverse set of measures: dietary intake (Martin et al. 2009); monitoring of food purchases for allergies via the use of smartphones with barcode readers (Arens et al. 2008); consumption at the moment of the purchase (Hosoe 2005); recreational saltwater fishing catch (Baker and Oeschger 2009); energy expenditures via embedded accelerometers (Lau et al. 2010); remote health care monitoring (Jea et al. 2008); Portable People Meters for the passive capture of radio program listening (Patchen and Webb 2002); and smart pill bottles to monitor prescription medicine compliance (Becker et al. 2009). The U.S. sample of the Homescan project utilizes a Universal Product Code scanner which facilitates electronic capture and transmission of consumer data on an on-going basis (Link 2010).

Concurrent with the changing data collection landscape, the environment for consumers has radically changed in the last 25 years. The increasing use of e-commerce[6] and the proliferation of "big box" stores have significantly impacted purchasing behavior. Product-based diaries seem antiquated when the box that arrives from Amazon.com or the shopping bags brought home from Sam's Club include products ranging from food to computers.

A wide array of technological tools and methodological innovations can be brought to bear on a redesign of the CE surveys. Technology may reduce underreporting while also reducing burden, facilitating longer retention as diarists and/or integrating the CED and CEI components within a single data collection, an approach used for the United Kingdom's Family Expenditure Survey (Battisin

---

that they felt the individual diaries were redundant and unnecessary since the unit makes purchases collectively.
[6] The Wall Street Journal estimates that e-commerce accounted for approximately 7% of U.S. shopping in 2010, excluding auto, travel and prescription drugs, with an annual growth rate of approximately 11%.

2003).[7]

Flexible Mode of Data Collection

Ecological validity is achieved in surveys when questions and response options are sufficiently flexible so as to be asked and answered in a manner that resonates with the respondent and his or her situation (Schaeffer 1991). The same logic can also be applied to the overall design of a data collection effort – the goal is to provide sufficient flexibility so as to achieve ecological validity. This approach has not received extensive treatment in the methodological literature, except for responsive survey designs (Groves and Heeringa 2006), most often implemented to control errors arising from nonresponse. We applaud BLS's initiative to move forward in consideration of flexible survey designs for the CE; clearly the nature of the task differs across different consumer unit compositions. Consumer units structured so as to have a central expenditure gatekeeper may require a very different approach to data collection than households with three adolescents and parents who each maintain their own credit cards and checking accounts. Thorough understanding of how various consumer units comply with the data collection requirements for the CEI and CED will be key to the development of alternative (flexible) data collection methodologies.

B.   Approach to CE Redesign

In the sections that follow, we outline our approach to the redesign of the Consumer Expenditure Survey. In making our recommendation, we attempt to address some of the key features of interest to the Bureau of Labor Statistics. To the extent possible, our recommendations draw from empirical literature or current experience and attempt to extrapolate to a design that would be viable in the future—thinking along a time frame beginning five or more years from 2011. We have attempted to ground our recommendations in theoretical and empirical literatures from a broad spectrum of disciplines that encompass data collection,

---

[7] There is some evidence that the integrated approach to the collection of consumer expenditures leads to aggregated values that more closely align with national accounts estimates than designs which involve separate samples for the interview and diary portions of the expenditure data collection (Banks and Johnson 1998).

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

(e.g., epidemiology, economics, sociology, political science); however, since much of the peer-reviewed literature lags significantly behind innovation, some of our recommendations are based on personal communications with respect to current applications and field operations incorporating technology, others are speculative in nature, based on our assessment of innovations currently being tested and the direction of technology in the near future.

As outlined in our technical proposal, we have attempted to design a consumer expenditure survey that is *prospective, flexible,* and *technological,* while still maintaining a probability sample of the population.

The report is outlined as follows:

- Section II provides a broad overview of the design recommendations;
- Section III provides details of the design incorporating research that supports the recommendations;
- Section IV outlines methodological research to support the recommended redesign;
- In light of our concern that the CE redesign address both measurement and nonresponse as sources of error that contribute to discrepancies between CE data and benchmarks, Section V discusses design features that are targeted at reducing nonresponse rates and potentially nonresponse error;
- Section VI discusses issues related to costs; and
- Section VII in which we discuss a broader range of alternatives for NAS and CE to consider.

II.   Summary of Recommended Design Features

A. Introduction

The challenges facing the Consumer Expenditure (CE) survey are numerous but not dissimilar to other large complex data collection efforts. These challenges include, but are certainly not limited to, pressures from a diverse set of data users for timely, detailed, and integrated data, response rates that continue to decline even for the most prestigious Federal data collection efforts, and increasing diversity with respect to both the products and services available to consumers as well as how and where those services and products are acquired. In a rapidly changing environment, that is, both the environment related to data collection as well as the consumer environment, how best to position the CE for the future?

We see our task not as one of suggesting tweaks to the questionnaire or modest recommendations; several panels have been convened in the past to offer recommendations along those lines (e.g., Mathiowetz 1987; Edgar 2011). Rather we view our task as the opportunity to step back, assess the data collection as well as consumer environment of the future, and offer a design that approaches the CE from a fresh perspective, one that harnesses technology so as to improve both participation and the quality of data and which provides flexibility in the approach to measurement, reflective of the trend in survey methodology for responsive and/or flexible approaches to design.

This section provides an overview of our recommended design for the CE. Subsequent chapters provide the theoretical and empirical support for our recommendations as well as suggestions for methodological research for those design recommendations for which the current literature provides little or conflicting data.

B. Design Recommendations

Sample Design

One of the clear requests from the CPI and outside data users is the desire of a single sample that integrates diary and retrospective reports of expenditures for consumer units (Edgar 2005; Henderson et al. 2011). We recommend moving to such a design, similar to the sample designs currently implemented in the United Kingdom and Canada (Horsfield 2010; Tremblay, Lynch and Dubreuil 2011). To meet the challenges of minimizing respondent burden while also providing data users with the ability to examine change within consumer units, we are recommending two components of the single-sample design: (a) a cross-sectional component, for which a CU would participate for a single month; and (b) a panel component, for which a CU would participate three times as a diarist – in month 1, month 7, and month 13— providing a longitudinal evaluation of spending while reducing burden across the full sample.

The design represents a radical departure from the current approach to collecting consumer expenditure data in several ways. First, the design minimizes reliance on retrospective recall. As is outlined below, the only expenditure items for which the respondent is asked to rely on retrospective recall are global questions and large expenditure items. Second, the reporting task is the same for the cross sectional and longitudinal components – specifically, maintaining a diary for a one month period. No one CU would provide data for a full quarter nor consecutively for a full year; yet the panel component facilitates within-CU change with the proposed design to examine change within an entire month (two more weeks than the current diary design permits), change for the CU from that month and six months from the initial month, and a comparison of expenditures for that initial month in year *t* and year *t+1*.

Moving to a single, integrated design provides flexibility in index construction for the CPI (Casey 2010), offers microdata users the advantage of accessing a single data source for the best estimate of expenditures, and

reduces redundancies that currently exist in the Consumer Expenditure Quarterly and Diary Surveys (CEI and CED). The challenges for such a design include addressing perceptions of and actual respondent burden for the individual sampled consumer units and shifting approaches to quarterly and annual estimation.

Questionnaire

Regardless of whether a CU was a sampled for the cross sectional component or the panel component, the data collection would involve the following:

- An initial face to face interview, in which basic demographic and socioeconomic information would be collected about the CU, including the administration of global questions concerning expenditures and income in the past year, itemization of items paid on a monthly basis (e.g., utilities) and automatic payments[8] and respondent training;
- Thirty day collection period in which CU members would record all purchases, expenditures, and payments made during that time;
- Use of technology-based or paper-based memory triggers for capturing expenditures in real time; and
- A face to face wrap up interview, during which the interviewer would verify expenditures, record additional expenditures or receipts for which the CU did not or could not record during the reference period (including those that may not have been captured in the diary due to automatic payments), and retrieve the technological devices used for data capture.

For CUs that comprise the panel component, we recommend a face to face interview prior to the resumption of serving as a diarist that would involve an update of demographic, socioeconomic, and consumer unit information and remind respondents on how to use the technology for data capture.

---

[8] Note that we are not recommending the collection of the expenditure amount in this first interview.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

Proactive and Flexible Data Collection

According to the original request for proposals issued by the National Academy of Sciences, the recommended redesign is intended to address (a) underreporting of expenditures in the household surveys; (b) changes in the survey environment – both the environment of consumer purchasing and the data collection environment; and (c) the need for increased flexibility in approaches in data collection. With these requirements in mind, the design we recommend to the National Academy of Sciences and to the Bureau of Labor Statistics incorporates the following:

- Use of computer tablets as the primary "diary" recording device for each sampled consumer unit;
- Web-based[9] data collection that facilitates reporting by individual diarists within the consumer unit;
- Capture and recording of both paper and electronic receipts;
- Portable paper diaries, cell phones, and smartphones for "real time" data capture to serve as memory triggers;
- Flexible data collection approaches that take advantage of CU's use of financial management software, credit card summaries, and/or checking account summaries; and
- CU-based respondent and technology design options.

We briefly explain each of these design features below, with further explication offered in Section III.

Our design addresses the issue of underreporting by minimizing reliance on retrospective reporting, promoting "real time" recording of all expenditures and payments, and emphasizing self reporting among all CU members. The use of a web-based diary, via web-enabled tablets, provides an efficient means by which

---

[9] For those CUs for which internet access is not possible, the computer tablet would include a computer-assisted questionnaire that mimics the web-based questionnaire. The tablet would serve as the data collection and storage device for these CUs, once again, with individual IDs for CU members to facilitate individual diarists within the CU.

each member of the CU can log on to his or her own personal diary to record expenditures.[10] The flexibility and computing power of a tablet will allow CE staff to develop an instrument that minimizes burden (e.g., pick lists; scanning of receipts and barcodes; ease of selecting repeat purchase items) and facilitates consistency in reporting at the level of detail necessary for the CPI. We envision a data collection approach with the tablet that allows for the use of apps, integration with other technology, online help for the CU members, and real time monitoring of diary entries by the CU.

Our recommended design attempts to reduce measurement error by encouraging real time data recording via self report, to the extent possible. Does this mean that every CU should utilize a tablet? Should we require all CU members to participate, even if one member of the CU indicates that he or she will serve as the CU reporter? With respect to technology, we encourage BLS to adopt a standard platform that incorporates technology that could be harnessed to improve reporting and reduce burden. Recent experience by Nielsen suggests that those unfamiliar with technology can be trained to use the technology for data capture (Link, et al. 2011). As the technology will be more than a method of accessing the survey, but will also incorporate scanning of receipts, barcode reading, QR reading, and other such advances, it makes sense to provide each household with a tablet rather than relying on the CU's own technology.

With respect to multiple reporters per CU, the limited literature suggests that the use of multiple diaries per CU increases the reporting of expenditure items and CU expenditures (Grootaert 1986; Edgar, et al. 2006). If the source of the increasing discrepancy between CE and the Personal Consumption Expenditure data from the National Accounts is due to measurement error, then increasing self reports and minimizing recall periods are two well established

---

[10] Once again, we note that for those who live in a part of the country without cellular phone service, the tablet would serve as the data collection device, with a computer-assisted interview questionnaire that mimics the web-based instrument. The major disadvantage will be that CE staff will not be able to monitor compliance in the same manner as is possible with a web-based instrument.

means for improving data quality (Bound, Brown and Mathiowetz 2001).[11] Furthermore, the use of technology, in which each member of the CU can log in to his or her individual diary with their own login and password, permits persons who make purchases that they would rather not have other members know about to answer confidentially (e.g., teenagers not wanting their parents to know about certain purchases), more so than if a paper diary is used (e.g., Stinson, To and Davis 2003).

The use of a computer tablet will facilitate the integration of scanner and barcode technology as part of the data capture. Obviously, in 2011, we are not in a position to recommend specific technology that will be available in five years. However, we have tested the feasibility of our proposed design using currently available hardware and software applications. Our feasibility test – described in detail in Section III – was completed using an iPad 2, an iPod Touch, several relatively inexpensive apps, and free financial management software.

"Receipts" involve far more than the saving of cash register receipts; although consumers may still receive receipts for some purchases (e.g., groceries), they are just as likely to receive receipts via email (for both brick and mortar purchases as well as internet-based purchases). For example, clothing stores such as Nordstrom and Banana Republic now regularly ask customers if they would like a paper receipt or have the receipt emailed to them. Similarly, bills may exist in paper format, as an electronic email, or simply be automatically deducted from a checking account or paid via automatic payment via credit card. For example, the 2009 Survey of Consumer Payment Choice reports that 65.7% of consumers have access to online banking and 10.1% access banking through their mobile device (Table 1, Foster et al 2011). Additionally, 48.8% of consumers currently use online banking and 56.3% have bills automatically deducted from their bank accounts (Table 4, Foster et al. 2011). To facilitate ease of capturing electronic receipts, we recommend the use of a personalized

---

[11]Edgar, et al, (2006) did find that some CU members felt the individual diaries were redundant and unnecessary since the unit makes purchases collectively. The resolution between these findings and are recommendations may be to develop a protocol that allows for some interviewer flexibility, in light of CUs asserting or requesting a single person serve as the respondent.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

email account or "file cabinet" to which CU members can forward receipts and link these electronic receipts with diary records based on the date of purchase. This "file cabinet" can be used by the respondent or the interviewer as a recall aid without the CU needing to have a printer, paper, or ink to access these records.[12]

Coupled with the technological approach offered by the use of computer tablets, we also encourage adoption of multiple portable means for capturing data that can serve as a memory trigger for CU members. Here we suggest a protocol in which the interviewers encourage each CU member to note purchases and save receipts, either through the use of their own smartphones, cellphones,[13] or simple paper pocket diaries and holders.[14] This is a dramatic expansion on the current CED recommendations to diary CUs to keep paper records in an attached 'pocket.' The use of these portable recording devices --- encompassing pictures, voice recordings, and/or old fashioned pen and paper – would be designed to serve as memory aides with the details recorded upon return to the home. Again, these could be linked to the diary based on the date of the purchase. This strategy recognizes the current findings of improved data quality with use of records in the CE (Safir and Goldenberg 2008) and builds from it by expanding records to incorporate other recall aids. Once again, the goal would be for each CU member to have a portable means for recording, maximizing technology already owned by CU members and for which they are comfortable. Included in this use of "portable" diaries are children, ages 7 to 15 who will be asked to maintain a simple paper diary.[15]

We note that CU members may already track finances or part of their household expenditures through the use of financial management software (such as those offered by mint.com, Quicken, Dave Ramsey, or Suze Orman), credit

---

[12] Several financial software packages as well as apps have a cloud based capability for tracking expenses and uploading receipts. For example ProOnGo (www.proongo.com) allows users to email receipts to a secure web portal.

[13] Coupled with the use of an inbound IVR data collection or simply a means for recording notes by the respondent.

[14] The use of an iPod Touch (or similar portable technology) for use as a scanner would provide the CU with an additional device for the collection of point of purchase data.

[15] A simple electronic "diary" could also be developed for those ages 7 to 15.

card expenditure category summaries, and/or checking account category monitoring. An obvious advantage to reducing respondent burden is to take advantage of recording practices already established within the consumer unit. In addition, financial management software such as Mint.com offer point of purchase applications to use with smartphones; as with pictures, voice recordings, and paper and pencil, we suggest leveraging these existing technologies as recall aids for those CUs that already utilize such software.[16] The integration of financial software is an area for which we recommend further exploration.

Field Protocol

To encourage participation and to facilitate training of CU respondents, we recommend that CE continue to recruit CU sample members via face to face contact.

A critical design issue is the length of the panel – that is, for how many weeks or months we ask CU respondents to serve as diarists. This is definitely an issue of cost-error tradeoffs, one that impacts the costs of data collection, the willingness to participate, the extent to which the data are impacted by panel conditioning/fall-off in reporting, and the need for month-to-month and/or year-to-year comparisons among the same CUs. No single design can optimize for all of these objectives, which is why we are recommending both a cross sectional and a panel component to the single integrated sample approach. We discuss the cost-error tradeoffs and address the issue of additional diary burden in detail in Section III. Details concerning the field protocol are also provided in Section III.

Our design recommendations also echo the sentiments expressed by Michael Link at the Workshop of Data Producers (June 1-2, 2011). Dr. Link's recommendations included engaging the respondent in a process that could be described as an interchange of information between the respondent(s) and the data collector. As a means for sustaining interest and to improve compliance,

---

[16] One could envision two approaches with respect to the use of financial management software: (1) a partnership with a particular company, so as to take advantage of already existing capabilities; or (2) developing CE-specific interfaces for the primary software packages used by consumers.

UWM UNIVERSITY of WISCONSIN MILWAUKEE   Abt SRBI   N

we recommend that BLS in its redesign of the CE develop feedback mechanisms to the CU. For example, an analysis of CU expenditures (e.g., percent distribution of expenditures) may provide the CU with beneficial feedback as well as serve to remind them to continue to record information.

A radical change in design comes with risks. Although we have reviewed the literature as well as current practices in recommending our design (and these are cited in the detailed Section III), we strongly recommend a number of methodological experiments to support our recommendation.

C.   Alternative Data Sources

Federal Data Sources

We consider the use of alternative data sources, primarily with respect to their use for benchmarking. At least three Federal surveys provide detailed information that could be (or has been in the past) utilized by BLS as benchmark data. The three surveys are the Medical Expenditure Panel Survey (MEPS), the National Household Food Acquisition and Purchase Survey (FOODAPS), and the Residential Energy Consumption Survey (RECS). Each of these surveys collects detailed information at the dwelling unit or household unit level with respect to expenditures for medical expenses, food, and energy use, respectively. Use of any of these surveys involves surmounting issues of definitional compatibility, reference period correspondence, and data sharing across Federal agencies.

Non-Federal Data Sources

Beyond the scope of data collected by the Federal government, there is an abundance of consumer data collected and warehoused in the private sector. How can the Federal government harness these rich data sources? Are there means for using these data to reduce respondent burden for CE sample members? Can these data be used as benchmarks, for example, within particular expenditure categories? Or can these data provide a means for improving estimation, once again, within particular expenditure categories? We were initially optimistic about micro-level integration of non-federal administrative data sources with CE data. However, the current state of knowledge about these

UWM
UNIVERSITY of WISCONSIN **MILWAUKEE**

Abt SRBI

N

sources and the incredible task involved in turning administrative records from private companies into survey data for all sampled persons makes us cautious in recommending their use for purposes other than nonresponse monitoring and benchmarks. This is, however, one area for which a recommendation made in 2011 may be outdated by 2015.

So in answer to the questions raised above, yes, we do believe that administrative data *could* be used to reduce burden but for which CUs and for what expenditures, it is difficult to determine. For example, with respect to retail grocery sales, in 2009 four companies (Walmart, Kroger, Costco, and Supervalue) account for more than 50% of sales in the U.S (*Supermarket News's* Top 75 Retailers for 2009, undated) and in the largest 100 metropolitan areas, these retailers accounted for 72% of sales by 1998 (Martinez 2007). The ramifications of this level of concentration is obvious for capturing consumer expenditures of groceries: for those consumer units who are enrolled in customer loyalty programs, one could capture these data from the provider rather than the consumer *if* BLS can successful hurdle the negotiations for micro level data from these organization AND can obtain permission to access micro data for individual consumer units.

But the downfall of the use of administrative records as outlined above is that no one set of data are universal, leading to mixed "modes" of data across consumer units or even individuals within the consumer unit.

With respect to use for benchmarking, once again, there are categories of expenditures and particular subgroups of the population for which these administrative records *could* be (and have in the past been) informative (e.g., the use of state tax records to benchmark expenditures on cigarettes, alcohol, and gasoline).

D. Addressing Issues of Nonresponse

Although issues of nonresponse were not highlighted as a major focus of design recommendations outlined in the NAS RFP, we felt we would be remiss to

limit our consideration of sources of error that potentially impact the quality of CE estimates to measurement. Comparisons between estimates based on the CE and those that are based on the National Accounts suggest deterioration in the quality of the CE data. While measurement error is one source that most likely contributes to the disparity in these two sets of estimates, other sources such as coverage error, differential post survey adjustment procedures, definitional differences, as well as nonresponse need to be considered. Olson (2011) found that measurement error in the CE was strongly associated with participation propensities (more so than with spending in various expenditure categories) and recommended that targeting design features to improve measurement error may also be beneficial to participation propensities. As a proactive means of addressing nonresponse, we recommend the integration of global expenditure questions at the time of the initial CE interview (conducted by an interviewer). A comparison of estimates based on global questions included in the Panel Study of Income Dynamics to CE expenditures obtained from the interview portion of the CE found that a limited number of global questions accounted over 70% of CE reported total expenditures and fell within the 95% confidence interval for CE across the time period of the study (Li et al. 2010). Responses to global questions could be used in two ways: (a) as a means for prompting respondents when reported expenditures fall below the annual global questions for a category and (b) as further information to be considered in nonresponse adjustment. Note that we make this recommendation for monitoring of potential nonresponse error, not as a means of addressing measurement error (following concerns voiced at the CE Methods Workshop, Edgar 2011).

Detailed discussion of considerations with respect to nonresponse can be found in Section V.

E. Costs

As part of the RFP, we were asked to discuss cost estimates related to development, implementation, and ongoing data collection. To the extent possible, these issues are outlined in Section VI. Our design attempts to be cost

neutral with respect to data collection and to take advantage, to the extent possible of existing technology (as opposed to reinventing the wheel).

III.   Recommended Design Features: Details

A. Sample Design

Cross Section and Panel Component

One requirement of the CE is to support micro-level analyses over a 12-month period (Henderson et al. 2011). This feature is not required for the CPI, but it does allow microdata users to track spending for a CU over the course of an entire year. The current design satisfies this panel requirement by conducting the quarterly interview for five quarters with each household, relying extensively on retrospective recall of expenditures. This protocol places substantial burden on participating households. The burden of completing the quarterly interview five times has significant implications for both nonresponse and measurement error, including the burden of collecting data not used in estimation. Households that attrite (do not complete all of interviews 2 through 5) have different characteristics than households completing all five interviews (King, et al. 2009) increasing the risk of nonresponse bias. In comparing attritors versus non-attritors, King and her colleagues found statistically significant differences on age, marital status, CU size, and region, race, educational attainment, and housing tenure. They concluded that the CE data "may not be missing completely at random." Furthermore, repeated interviewing may introduce panel conditioning problems, whereby respondents are motivated to misreport in order to shorten the length of the interview (Safir 2011).

In the interest of reducing respondent burden – and in turn reducing the joint risks of nonresponse and measurement error – we recommend a design in which only a subset of CE households are empanelled. The panel component would be comprised of a sub-sample of CUs that are interviewed three times at six month intervals. The CU roster and CU socioeconomic information, including income, would be updated for each interview. This component would be staggered such that a random 1/12 of the yearly sample (and a random 1/3 of the quarterly sample) is interviewed for the first time each month. Households

selected for the larger cross-section component, by contrast, would only be included for only a single month, with face to face interviews prior to and after the diary month.

This design permits within-CU comparisons to examine change within a single month, two more weeks than the current CED design. It permits direct within-CU comparisons for one year change estimates across two consecutive years (e.g., January 2012 and January 2013) while also providing data for the same CU at the midpoint of panel period (in this case, July 2012), facilitating six-month change estimates. The minimum needed to satisfy the requirement of providing longitudinal data for micro-level analyses is two interviews conducted 12 months apart. We realize of course, that although the proposed design does not provide data for a CU for each week, and thus all twelve months, it does facilitate evaluating year to year changes in consumption by a consumer unit. This is a critical element of our design. Our design relies almost exclusively on "prospective" and real time measurement of expenditures; the exception being global recall questions asked primarily for purposes of nonresponse bias analyses and adjustment and salient, high expenditure items (see details below). Regardless of whether a CU is included in the cross-sectional sample or the empanelled sample, the respondent task is the same – complete the diary for a one-month period. The difference lies in the repeated request to the empanelled members –with the collection of the one-month diary for months 1, 7 and 13.

Our recommendation is based on the requirement that the proposed design *not* rely on retrospective recall for expenditure estimates. In this design, we echo concerns voiced by experts from the CE Methods Panel held by the Census Bureau and COPAFS, including Gordon Willis's statement "…people can't recall certain things….I mean, there are no sophisticated magical procedures that we can necessarily use, at least at reasonable cost here" and Peter Miller's urging to "Just stop asking questions and get them to do the behaviors so that you'll have the information you want, and it will be better information." Could CUs be empanelled for a full quarter or a full year? That, of course, is an empirical question that could be tested. We do not recommend such a design that would

UWM
UNIVERSITY of WISCONSIN **MILWAUKEE**

Abt SRBI

N

require serving as a diarist for 12 months or 52 consecutive weeks as we believe this would be unduly burdensome to respondents and subject to high levels of respondent fatigue. We could envision a number of alternative designs to meet the needs of data users who require full quarter or annual expenditures for a single CU. All of those designs, however, involve some use of retrospective recall. As noted in the RFP, "[A] recall survey would only be used to supplement information not otherwise available" (p. 13). So to meet the needs of these outside users, BLS could implement one of the following supplemental designs:

- A panel using the current CEI design;
- A panel with monthly retrospective interviews (thereby reducing the length of the recall period) but increasing respondent burden;
- Incorporating retrospective questions for the empanelled members of the proposed design; or
- Statistical imputation of the in-between months for the panel members, borrowing information and strength from the cross-sectional units and other panel members to improve the imputation models.

Under the assumption that the prospective, diary approach yields better quality data, estimates from any or all of these designs could be adjusted with the monthly diary data. But each of these alternatives is simply a tweaking (or maintenance) of the status quo and therefore, not recommended for meeting the primary objectives of improving data for the CPI.

The recommendation to interview the empanelled CUs three times is based on considerations of respondent burden, measurement, and sample tracking costs. Minimizing respondent burden is a key priority, which means that a lower number of interviews are desirable. The approach of conducting just two interviews spaced by 12 months, however, is in our view suboptimal for several reasons. A full year is a significant length of time during which we would expect the composition of many CUs to change and/or the entire CU to move to a new address. With respect to measurement, having that long of a gap between data collection periods may also require a fair amount of interviewer "re-teaching" at

interview 2 and due to CU members having forgotten what they learned at interview 1 about how to document and report their expenditures. Adding an additional interview six months after the first interview is intended to alleviate these concerns. There will still be some churn in CU composition and forgetting of the reporting protocols, but the magnitude of these issues, we suspect, would be less severe at six month intervals rather than a 12 month interval.

The recommendation to move away from a fully empanelled interview sample is one of the most significant features of this proposed redesign. This is a major opportunity to reduce the burden on the majority of households participating in the CE and, in turn, improve data quality. Importantly, because all expenditures will be collected prospectively through the diary, there will be no need to discard the first interview as in the current design. The fact that a panel design is not required by the CPI is a critical consideration in this recommendation.[17] While the needs of other data users are important, the tremendous burden on each cooperating CU inherent in a five interview design or potential measurement errors induced through extensive use of retrospective recall does not appear to be justified. Instead, we recommend that the CE support longitudinal micro-level analysis by collecting these data on just the panel component.

We expect greater reporting accuracy under the redesign because the empanelled households will be better compensated and subjected to fewer interviews, reducing the likelihood of errors caused by attrition and panel conditioning. In addition, a major benefit of the panel component is that it will collect all expenditures at the micro level for three full months, which is an improvement over the current design that does not provide both interview and diary data at the micro (CU) level.

---

[17] We are cognizant that other data users do make use of total quarterly expenditures and spending across all four quarters (Henderson, et al, 2010). Any redesign must face the competing constraints of meeting the needs of the CPI, reducing measurement and nonresponse error, minimizing respondent burden while at the same time maximizing the utility of the data for the majority of data users. Our design attempts to meets these competing demands, with priority given to the CPI's data needs and the burden placed on the respondent.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

Sample Size and Statistical Precision

According to Henderson et al. (2011), the current CE sample size is not sufficient for supporting desired precision levels. The current sample size is 7,000 non-bounding interviews (interviews 2 through 5) per quarter in the Interview Survey, and 7,100 interviewed households per year in the Diary Survey. In order to achieve a coefficient of variation (*CV*) of one percent or less for the mean annual total expenditures per CU at the national level, it is estimated that the sample sizes would need to be at least 7,500 non-bounding interviews per quarter in the Interview Survey, and 7,500 interviewed households per year in the Diary Survey (Henderson et al. 2011).

Mapping these sample size requirements onto the proposed sample design is challenging for several reasons. One reason is simply the lack of information. We were unable to find any estimates of the average design effect associated with the multi-stage sample or the nonresponse and calibration weighting adjustments used for the current CE. This limits the utility of knowing the nominal samples reported in CE methods reports (e.g., Swanson 2011). A second challenge is that the proposed design differs from the current design in several important ways. Most notably, the proposed design features one survey with two samples (cross-sectional and panel) instead of two separate surveys (Diary and Quarterly Interview).

In calculating the statistical precision provided by the current design, it seems appropriate to only use one of the surveys (e.g., n=28,000 households interviews per year in the Quarterly Interview) because the other survey (Diary) collects, for the most part, a separate set of expenditures. Pooling the two survey samples would, thus, seem to yield overestimates of the precision provided by the current CE.

In determining the optimal sample size to propose for the redesign, we first sought to calculate the sample size needed to achieve the one percent or less *CV* for the mean annual total expenditures per CU at the national level mentioned in Henderson et al. (2011). Assuming the mean is approximately $49,067 (Swanson 2011), the standard error of the mean required to achieve

CV=1% would be approximately \$490.67. Next, we needed an estimate for the population variance. Using a standard error of the survey mean of \$595 (Swanson 2011) and an annual sample size of n=28,000 (Henderson et al. 2011) and assuming a design effect equal to 1.0, the estimated population standard deviation is 99,562.54. Finally, we can estimate the annual sample size need to achieve CV=1% as $(99,562.54/490.67)^2$ = 41,173. This would be 1.47 times more sample than is used under the current design.[18] Given that the proposed redesign should not dramatically increase the CE program budget, we simply note the disparity between the current sample size and what we estimate to be the sample size required to achieve the stated precision target. We do not use the annual n=41,173 figure elsewhere in this report because we suspect that it is not realistic under the program budget. Instead, the sample sizes we propose are designed to roughly approximate the statistical precision provided by the current CE interview sample sizes. If more funding becomes available, it would be fairly straightforward to increase the sample size.

For the proposed redesign, we recommend completing 7,000 household interviews per quarter in the cross-sectional component and 1,350 household interviews per quarter in the panel component. At the start of the second year of the redesigned CES, the 1,350 panel interviews would be the aggregation of approximately 450 1st interviews (e.g., 150 in each January, February and March), 450 2nd interviews, and 450 3rd interviews. Prior to the start of the second year, the panel component sample size would be somewhat smaller because the full rotating design is not yet in place.

The proposed design for the cross-sectional and panel components is illustrated in Figure 1. The panel component sample sizes in the figure do not account for attrition, mainly because the attrition rate for the proposed design is currently unknown. We would note, however, that we do expect the number of

---

[18] We note that as independent, outside researchers we do not have line of sight to all of the relevant design factors and estimated population parameters that are available to the CE sampling statisticians. The estimates reported here represent out attempt to compute the sample size needed to satisfy the stated precision requirements based on CE figures reported in the literature. We acknowledge that our estimates may somewhat imprecise due to the aforementioned lack of information about the relevant aspects of the current CE sample design, such as the design effect due to stratification, clustering or weighting..

UWM UNIVERSITY of WISCONSIN MILWAUKEE   Abt SRBI   N

completed CU interviews to decrease somewhat across the three interviews within a given cohort. Pilot testing is needed to determine exactly how much attrition should be expected between the first, second, and third interviews in the panel component, and thus how much the sample should be inflated to meet a third interview target sample size of 150 in the panel component. For example, if we assume a 70% cooperation rate to the first interview and an 80% cooperation rate to each of the second and third interviews, BLS would need to sample 335 cases each quarter to obtain 235 completes (335*0.70) in the first interview, 188 completes in the second interview (235*0.80), and 150 completes in the third interview (188*0.80).

While the cross-sectional component sample sizes are presented here on a quarterly basis, it is important to keep in mind that the period of data collection for each household is one month. Cross-section households and panel households would be instructed to report their expenditure for 30 days (not 90). Quarterly estimates of total expenditures would be made from adding across consecutive months.

Another consideration with respect to the CE sample size is the CPI requirement that Entry Level Items (ELI) be selected based on expenditure-based probabilities that are representative at the PSU level. CE does not currently meet this requirement and CPI must aggregate expenditures to the ELI-Region level in order to have a large enough sample for each probability (Casey 2010). ELI probability selection requires the largest expenditure sample sizes of any of the current CPI production uses because sufficient sample is required at the ELI-PSU level. Unfortunately, the data required to incorporate this CPI requirement into the sample design do not appear to be publically available or contained in any Gemini Project-related documents. While means, standard errors, and CVs are published in tables at the expenditure class level, we were unable to find such statistics at the ELI level, which are two levels below the expenditure class level.

	Year 1												Year 2												Year 3											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

Cross-sectional Component

	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec
Yr 1 samples	7,000	7,000	7,000	7,000								
Yr 2 samples					7,000	7,000	7,000	7,000				
Yr 3 samples									7,000	7,000	7,000	7,000

Panel Component

Yr 1 Jan Cohort — 150 (1st, Yr1 Jan), 150 (2nd, Yr1 Jul), 150 (3rd, Yr2 Jan)
Yr 1 Feb Cohort — 150 (1st, Yr1 Feb), 150 (2nd, Yr1 Aug), 150 (3rd, Yr2 Feb)
Yr 1 Mar Cohort — 150 (1st, Yr1 Mar), 150 (2nd, Yr1 Sep), 150 (3rd, Yr2 Mar)
Yr 1 Apr Cohort — 150 (1st, Yr1 Apr), 150 (2nd, Yr1 Oct), 150 (3rd, Yr2 Apr)
Yr 1 May Cohort — 150 (1st, Yr1 May), 150 (2nd, Yr1 Nov), 150 (3rd, Yr2 May)
Yr 1 Jun Cohort — 150 (1st, Yr1 Jun), 150 (2nd, Yr1 Dec), 150 (3rd, Yr2 Jun)
Yr 1 Jul Cohort — 150 (1st, Yr1 Jul), 150 (2nd, Yr2 Jan), 150 (3rd, Yr2 Jul)
Yr 1 Aug Cohort — 150 (1st, Yr1 Aug), 150 (2nd, Yr2 Feb), 150 (3rd, Yr2 Aug)
Yr 1 Sep Cohort — 150 (1st, Yr1 Sep), 150 (2nd, Yr2 Mar), 150 (3rd, Yr2 Sep)
Yr 1 Oct Cohort — 150 (1st, Yr1 Oct), 150 (2nd, Yr2 Apr), 150 (3rd, Yr2 Oct)
Yr 1 Nov Cohort — 150 (1st, Yr1 Nov), 150 (2nd, Yr2 May), 150 (3rd, Yr2 Nov)
Yr 1 Dec Cohort — 150 (1st, Yr1 Dec), 150 (2nd, Yr2 Jun), 150 (3rd, Yr2 Dec)
Yr 2 Jan Cohort — 150 (1st, Yr2 Jan), 150 (2nd, Yr2 Jul), 150 (3rd, Yr3 Jan)
Yr 2 Feb Cohort — 150 (1st, Yr2 Feb), 150 (2nd, Yr2 Aug), 150 (3rd, Yr3 Feb)
Yr 2 Mar Cohort — 150 (1st, Yr2 Mar), 150 (2nd, Yr2 Sep), 150 (3rd, Yr3 Mar)
Yr 2 Apr Cohort — 150 (1st, Yr2 Apr), 150 (2nd, Yr2 Oct), 150 (3rd, Yr3 Apr)
Yr 2 May Cohort — 150 (1st, Yr2 May), 150 (2nd, Yr2 Nov), 150 (3rd, Yr3 May)
Yr 2 Jun Cohort — 150 (1st, Yr2 Jun), 150 (2nd, Yr2 Dec), 150 (3rd, Yr3 Jun)
Yr 2 Jul Cohort — 150 (1st, Yr2 Jul), 150 (2nd, Yr3 Jan), 150 (3rd, Yr3 Jul)
Yr 2 Aug Cohort — 150 (1st, Yr2 Aug), 150 (2nd, Yr3 Feb), 150 (3rd, Yr3 Aug)
Yr 2 Sep Cohort — 150 (1st, Yr2 Sep), 150 (2nd, Yr3 Mar), 150 (3rd, Yr3 Sep)
Yr 2 Oct Cohort — 150 (1st, Yr2 Oct), 150 (2nd, Yr3 Apr), 150 (3rd, Yr3 Oct)
Yr 2 Nov Cohort — 150 (1st, Yr2 Nov), 150 (2nd, Yr3 May), 150 (3rd, Yr3 Nov)
Yr 2 Dec Cohort — 150 (1st, Yr2 Dec), 150 (2nd, Yr3 Jun), 150 (3rd, Yr3 Dec)
Yr 3 Jan Cohort — 150 (1st, Yr3 Jan), 150 (2nd, Yr3 Jul)
Yr 3 Feb Cohort — 150 (1st, Yr3 Feb), 150 (2nd, Yr3 Aug)
Yr 3 Mar Cohort — 150 (1st, Yr3 Mar), 150 (2nd, Yr3 Sep)
Yr 3 Apr Cohort — 150 (1st, Yr3 Apr), 150 (2nd, Yr3 Oct)
Yr 3 May Cohort — 150 (1st, Yr3 May), 150 (2nd, Yr3 Nov)
Yr 3 Jun Cohort — 150 (1st, Yr3 Jun), 150 (2nd, Yr3 Dec)
Yr 3 Jul Cohort — 150 (1st, Yr3 Jul)
Yr 3 Aug Cohort — 150 (1st, Yr3 Aug)
Yr 3 Sep Cohort — 150 (1st, Yr3 Sep)
Yr 3 Oct Cohort — 150 (1st, Yr3 Oct)
Yr 3 Nov Cohort — 150 (1st, Yr3 Nov)
Yr 3 Dec Cohort — 150 (1st, Yr3 Dec)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Total Interviews	7,450	7,450	7,900	7,900	8,350	8,350	8,350	8,350	8,350	8,350	8,350	8,350

Key
- 1st interview
- 2nd interview
- 3rd interview

Figure 1. Effective sample sizes for the cross-sectional and panel components under the proposed redesign

Absent the relevant data, we cannot make definitive statements about whether or not the proposed redesign would satisfy this requirement. The proposed design calls for 33,220 total interviews per year; this compares to the current data collection for the CEI of 35,000 completed interviews with estimates based on 28,000 interviews. This increase would be expected to bring this CPI goal more within reach, especially for the larger PSUs.

Stratification and allocation

The CE data users indicate a variety of needs not currently met in the current sample design, including greater desired precision from estimates on rural CUs and on military CUs as well as state level estimates. Furthermore, data users want more households at each end of the income distribution. We have insufficient information for examining tradeoffs between different stratification and allocation schemes. We simply note that other federal surveys that support state-level analyses at a monthly basis have samples that are substantially larger than the CE's current design (e.g., the CPS selects 72,000 housing units from 824 sample areas to obtain 60,000 occupied households; BLS, Technical Notes, pp. 190-192).

Weighting Protocol

The weighting protocol needed for the proposed redesign would have many similarities to the protocol used for the current CE design (Swanson 2011). Under the current protocol, each CU is assigned a weight, which is the number of similar CUs in the U.S. civilian noninstitutionalized population that the sampled CU represents. The final weight is the product of four components:

Final Weight = Base Weight x Weighting Control Factor x Nonresponse
Adjustment Factor x Calibration Adjustment Factor

The base weight is equal to the inverse of the CU's probability of being selected for the sample. Base weights in the CES are typically around 10,000. The weighting control factor adjusts for subsampling in the field. Subsampling occurs when an interviewer visits a particular address and discovers multiple housing units where only one housing unit was expected.

The nonresponse adjustment factor adjusts for interviews that cannot be conducted due to refusal or the inability to contact anyone at the sample unit. This adjustment is computed separately for each of 64 demographic groups defined by region, CU size, tenure (own/rent), and race (black/non-black). Within each group, the adjustment is computed as the ratio of the number of occupied housing units to the number of responders. As part of the CE redesign, we recommend investigating a more elaborate protocol for nonresponse adjustment. In particular, Hispanic ethnicity, education, presence or absence of children, and household type (e.g., married couple, single adult, nonfamily) are all robust correlates of survey nonresponse, and some characteristics can be reliably observed by interviewers (see Nonresponse section below). It seems possible that incorporating these other dimensions into the nonresponse adjustment may yield nonresponse bias reduction beyond that achieved under the current protocol. Of course, the use of additional variables may mean that a response propensity based adjustment may be needed rather than a weighting class adjustment. This change in methods for sample-based nonresponse adjustments may be especially warranted to the extent that auxiliary administrative data sets are identified that contain adjustment variables (again, see Nonresponse section).

The calibration adjustment adjusts the weights to 24 "known" population counts to account for frame undercoverage. The population counts are derived from the Current Population Survey and account for age, race, household tenure (owner/renter), region of the country, and urban/rural. Each CU is given its own unique calibration factor. There are infinitely many sets of calibration factors that make the weights add up to the 24 population counts, and the Consumer Expenditure Survey selects the set that minimizes the amount of change made to the "initial weights." The calibration adjustment factor is computed as the ratio of the CPS population estimate to the CE population estimate. The CE population estimate is computed as the product of the number of responders, number of people per CU and the nonresponse adjustment factor.

UWM
UNIVERSITY of WISCONSIN **MILWAUKEE**

Abt SRBI

N

The major changes proposed for the CE in this report focus mainly on modes of data collection and measurement. They are not focused on how sample units are selected from the sampling frame. In fact, we do not have access to the kinds of information used to evaluate different sample methods such as the work by Rawlings, et al. (2010). Consequently, we believe that the steps involved in weighting the redesigned CE would remain largely unchanged. That said, the distribution of base weight factors and other adjustments would be expected to change. For example, if the proposed sample sizes for the redesign are adopted (Figure 1), then the mean base weight value would be expected to be decrease slightly. Given the reduced respondent burden and focus on technology under the proposed redesign, we would also anticipate improved survey participation – particularly among young adults and perhaps other groups that traditionally exhibit lower survey response propensities. If this is borne out, then the nonresponse adjustments needed under the redesign would tend to be smaller than those required under the current protocol.

B. Survey Design and Field Protocol

Measurement error increases with the length of time between a transaction and the reporting of that transaction (Eisenhower, Mathiowetz and Morganstein 1991). The current design of the Quarterly Interview, with its emphasis on retrospective episodic recall, contributes to measurement error. These measurement errors are most likely in the direction of errors of omission as the discarded first wave bounding interviews are intended to limit telescoping. The effect of long recall periods and reliance on retrospective episodic recall are further exacerbated by lack of knowledge on the part of proxy reporters (e.g., Blair et al. 1991). With these error sources in mind, our design attempts to maximize the "real time" collection of expenditures and payments, with data provided to the extent possible by the person making the expenditure or payment.

The key components of our survey design involve:

- An initial face to face interview during which data are collected on the consumer units' demographic characteristics (including income), technology ownership and use,[19] global questions concerning expenditures in the past year, questions which assess expenditure monitoring by the CU, and retrospective questions concerning the purchase of low frequency, high cost items (e.g., automobiles) in the past year;

- Training during the initial interview (and during follow up visits) as to what and how to record information;

- The placement of an electronic diary recording device (e.g. tablet) that encompasses a web-based as well as a computer assisted diary with individual IDs for each member of the CU ages 16 and older and which would be maintained for a 30 day diary period;

- Placement of additional hardware to facilitate the scanning of barcodes and the electronic capture of receipts, bills, and other documents;

- Placement of paper portable diaries (or training to use CU-owned electronic devices[20] to be used as portable diaries) for CU members ages 16 and older;

- Placement of paper mini-diaries for children ages 7 to 15 years of age[21];

- Electronic file cabinets for CU members to upload electronic receipts;

- Ongoing monitoring of compliance throughout the one-month diary period; and

---

[19] Knowing what technology is owned by CU members and how the various devices are used will be important in matching the various design features of the expenditure survey and diary to the technology level of the CU.

[20] If a device such as an iPod Touch is placed in the home for the purpose of barcode reading and /or scanning, that device could also be used by CU members for the purposes of point of purchase data capture.

[21] Or a simple electronic-based diary for this age group.

- Wrap up interview that encompasses scanning of receipts (as needed), review of expenditures, collection of payments that may not have been captured in the diary (e.g., automatic deductions from accounts or payroll), adjudication of items that may involve reimbursements (e.g. medical expenses) and retrospective questions as needed to establish as complete a record as possible for the CU.

Modes of Data Collection

In our proposal, we suggested that our final recommendation would advocate for the use of multiple modes of data collection for the diary, given recent empirical work that suggested that some respondents may be comfortable completing a diary online while others would prefer a paper-based diary (Olson, Smyth and Wood 2010). However, our review of the literature as well as our interviews with organizations that have adopted and integrated technology for data collection suggests that it is feasible to utilize technology in data collection across a broad range of demographic groups (personal communication with Nielsen staff, July, 2011; personal communication with RTI staff, July, 2011). For this reason, we are recommending the use of a computer tablet for all sampled CUs. Why the tablet? In its review of technology for CE data capture, Westat (2011) noted several desirable features of tablets: large storage capacity (up to 64 gigabytes), wireless connectivity, wide range of applications, the ability to easily run multiple applications (multitasking), and embedded GPS capabilities. Tablet technology also offers the advantage of developing applications that could, if desired, be ported to other more mobile devices. We envision CU respondents engaging in two very different and distinct types of data collection – (1) the collection of bits of information that can serve as memory triggers and (2) the actual CE diary. With respect to the former (discussed below) we do recommend a range of modes, drawing upon the level of technological expertise and ownership within the CU. However, for the actual diary, we see major advantages of a singular design and mode of data capture, including but not

limited to the cost savings associated with a single design and training of interviewers. Furthermore, use of an electronic instrument in which items can be automatically coded into relevant expenditure categories should decrease post-data collection processing time, and thus increase timeliness of the released data.

We also recommend harnessing the power of the tablet/web-based instrument for easing the respondent burden in reporting redundant purchases. For example, we speculate that a significant portion of grocery store purchases are redundant from week to week. The creation of dynamic smart lists – from which the respondent could select for subsequent diary entries – is but one feature that could reduce the day to day respondent burden.

In recommending a web-based data collection protocol, we recognize that although the majority of consumer units in the U.S. have the capacity to complete a diary online (Pew Research Center 2010), a small but significant portion (22%) of the households in the U.S. (in 2011) cannot access the internet from home. Furthermore, penetration of high speed internet access is even lower (about 66%) (Smith 2010). The use of a tablet-based CE diary allows the flexibility to allow those CUs with web access (or for whom web access is possible) to complete the diary via the web while offering those CUs without access the same instrument via a computer-assisted self-interview. The major advantage of such a design is that the use of a single technology and user-interface negates the confounding of mode effects (and corresponding measurement error issues) with demographic characteristics associated with web access and/or technology ownership. Recent experience (e.g. Link et al. 2011) suggests that users with little to no experience using "high tech" devices can be trained to use such devices effectively for data collection.

Survey Instruments

Our recommended design for the interview and diary draws upon the designs of the UK Expenditure and Food Survey (Horsfield 2010; Leicester and Oldefield 2009), the Nielsen Homescan project (Einav, Leibtag and Nevo 2008; personal communication), the experimentation related to the Canadian

household expenditure survey (Tremblay, et al. 2011) and the recent efforts related to the design of the USDA's National Household Food Acquisition and Purchase Survey (Matulewicz, Kalb, Redel and Cole 2011).

As with the current design, we recommend that the initial recruitment into the sample and interview be conducted face to face. We see no means by which one can efficiently recruit CU members and place the materials for a study such as this to serve as a responding household in a large, diverse national sample without a face to face contact. The initial interview would focus on establishing the membership of the CU, collecting demographic information (including income) about those CU members and collecting two sets of expenditures questions – retrospective questions focusing on major purchases and global expenditure questions.   In addition, the initial interview would identify key aspects concerning the CU's expenditure monitoring behavior (e.g., use of expenditure monitoring software; use of single credit cards for all purchases; check monitoring).

Where our recommendation deviates significantly from the current CE diary and from other national consumer expenditure surveys is with respect to the length of the diary period. Our recommendation is for diarists to serve for a one-month period, a period that doubles the length of the diary commitment for the current CE diary. Why the extended diary period? Clearly, there is experience in the private sector with households participating for extended periods of time recording expenditures. For example, Homescan respondents may participate for a period in excess of six months (personal communication); in the U.K., members of the UK-Worldpanel provide data for as many months as they wish to participate (Leicester and Oldfield 2009). Although a one month diary may seem more burdensome than a current two week design, the technology allowing for recognition over recall, use of 'pick lists' for previously purchased items, and electronic memory triggers should help reduce the cognitive burden on the CU even though the time burden will have increased.

Given the costs involved with training the consumer unit to use the tablet as a data collection instrument and the desire of data users to have fully

integrated information (that is, data that are currently collected in the diary and the interview) for CUs we believe that this design will be a reasonable balance of survey errors and survey costs. Unfortunately, the empirical literature to inform this design feature either does not exist or is more than four decades old (e.g., Sudman and Ferber 1971; Sudman and Bradburn 1973). This is a critical design decision and one which we strongly recommend further empirical work be conducted and for which we have outlined possible methodological investigations in Section IV.

Why the inclusion of the two sets of retrospective expenditure questions when our focus is on "real time" capture of expenditures? With respect to major purchases, we realize that the rarity of the event requires the reporting of these types of expenditures for an extended recall period. Without these retrospective reports, there may be insufficient data provided during the diary period for estimation of these rare, high cost expenditures. In addition our recommendation is based on the U.K experience (Horsfield 2010), coupled with the empirical data that suggests that major expenditures are less subject to errors of omission even for relatively long recall periods (Sudman and Ferber 1971; Neter and Waksberg 1964; Neter 1970). Although salient expenditures are less likely to be subject to errors of omission, the reporting of these expenditures may be more likely to be subject to forward telescoping, that is, the reporting of the purchase as occurring more recently than is true. Question wording for these retrospective reports will need to incorporate means to reduce forward telescoping for these items (e.g., two time period reports within a single interview). The U.K. questions focus on expenditures in excess of 5,000£; for the U.S. we would recommend an approach that focuses both on the level of expenditure and the type of expenditure (e.g., home repairs in excess of $5,000 made during the last six months; automobiles, regardless of price).

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

The inclusion of the global questions provides data that can facilitate both benchmarking as well as nonresponse adjustment.[22] As noted earlier, analysis of global questions included in the Panel Study of Income Dynamics (PSID) found that a set of global expenditure items accounted for 70% of the expenditures captured in the Consumer Expenditure Survey Interview data (Li et al. 2010).[23] For example, the PSID includes a global question concerning average weekly expenditures for food used at home. Such information could be informative with respect to understanding compliance and possible underreporting of food purchases during the diary period. Other global questions collected as part of the initial interview would allow analysts to examine the extent to which the diarists are reporting at a level commensurate with their past expenditure patterns and provide useful data for nonresponse bias detection and adjustment.

Table 1 summarizes the source for each category of expenditure currently collected in the CE and how the proposed design will collect this information. For some items, the source will be identical. For example, most food items are currently collected via the diary and the proposed design continues to have them collected through the diary. Similarly, we recommend that automobile purchases continue to be collected via the interview. However, other items, such as gasoline purchases are currently used from the interview. We believe that this type of smaller payment could be more accurately collected via the diary. Additionally, with an extended diary time of one month, this will be sufficient to capture

---

[22] Global questions could also be expanded to facilitate quarterly and/or annual estimation of expenditures for the empanelled members of the study. Although we don't recommend this approach, we realize that this compromise may be necessary to meet the demands of outside data users who rely on quarterly and annual estimates for individual CUs. However, such an approach simply reasserts the present design, with a mix of diary and retrospective reports, albeit for this case, for a single CU (as compared to the present CE design which draws from different CUs for the prospective and retrospective reports). We do, however, think that the global questions could provide useful covariates should imputation of between-wave expenditures be considered.

[23] Examples of these questions include "How much do you spend on food in an average week?; "How much are your monthly mortgage payments?" "How much is your total yearly homeowner's insurance premium?" "In the last month, how much did you and your family living there pay for each of these transportation related expenses....Gasoline? Repairs and maintenance? Parking and carpool? Bus fares and train fares? Taxicabs? Other transportation? See Exhibit 1, Li et al. 2010.

regular, consistent expenditures that occur monthly such as mortgage and rent payments and utility bills.

Table 1: Expenditure item collected in the Current Expenditure Survey by source, current design and proposed redesign

	Current Design	Proposed Redesign
Food		
Food at home		
Cereals and bakery products	Diary	Diary
Cereals and cereal products	Diary	Diary
Bakery products	Diary	Diary
Meats, poultry, fish, and eggs	Diary	Diary
Beef	Diary	Diary
Pork	Diary	Diary
Other meats	Diary	Diary
Poultry	Diary	Diary
Fish and seafood	Diary	Diary
Eggs	Diary	Diary
Dairy products	Diary	Diary
Fresh milk and cream	Diary	Diary
Other dairy products	Diary	Diary
Fruits and vegetables	Diary	Diary
Fresh fruits	Diary	Diary
Fresh vegetables	Diary	Diary
Processed fruits	Diary	Diary
Processed vegetables	Diary	Diary
Other food at home	Diary	Diary
Sugar and other sweets	Diary	Diary
Fats and oils	Diary	Diary
Miscellaneous foods	Diary	Diary
Nonalcoholic beverages	Diary	Diary
Food prepared by consumer unit on out-of-town trips	Interview	Diary
Food away from home	Diary + Interview	Diary
Alcoholic beverages	Diary + Interview	Diary
Housing		
Shelter		
Owned dwellings	Interview	Diary + Interview
Mortgage interest and charges	Interview	Diary + Interview
Property taxes	Interview	Diary + Interview
Maintenance, repairs, insurance, other expenses	Interview	Diary + Interview
Rented dwellings	Interview	Diary + Interview
Other lodging	Interview	Diary + Interview
Utilities, fuels, and public services		
Natural gas	Interview	Diary + Interview
Electricity	Interview	Diary + Interview
Fuel oil and other fuels	Interview	Diary + Interview
Telephone services	Interview	Diary + Interview
Water and other public services	Interview	Diary + Interview
Household operations		
Personal services	Interview	Diary
Other household expenses	Interview + Diary	Diary
Housekeeping supplies		
Laundry and cleaning supplies	Diary	Diary
Other household products	Diary	Diary
Postage and stationery	Diary	Diary
Households furnishings and equipment		
Household textiles	Diary + Interview	Diary

	Current Design	Proposed Redesign
Furniture	Interview + Diary	Diary + Interview
Floor coverings	Interview	Diary + Interview
Major appliances	Interview	Diary + Interview
Small appliances, misc. housewares	Interview + Diary	Diary
Misc. household equipment	Interview + Diary	Diary
Apparel and services		
Men and boys		
Men, 16 and over	Interview + Diary	Diary
Boys, 2 to 15	Interview + Diary	Diary
Women and girls		
Women, 16 and over	Interview + Diary	Diary
Girls, 2 to 15	Interview + Diary	Diary
Children under 2	Interview + Diary	Diary
Footwear	Diary	Diary
Other apparel products and services	Interview + Diary	Diary
Transportation		
Vehicle purchases (net outlay)		
Cars and trucks, new	Interview	Interview
Cars and trucks, used	Interview	Interview
Other vehicles	Interview	Interview
Gasoline and motor oil	Interview + Diary	Diary
Other vehicle expenses		Diary
Vehicle finance charges	Interview	Diary
Maintenance and repairs	Interview + Diary	Diary
Vehicle insurance	Diary	Diary
Vehicle rental, leases, licenses, and other charges	Interview	Diary
Public transportation	Interview + Diary	Diary
Health care		
Health insurance	Interview	Diary + Interview
Medical services	Interview	Diary + Interview
Drugs	Diary + Interview	Diary + Interview
Medical supplies	Interview + Diary	Diary
Entertainment		
Fees and admissions	Interview + Diary	Diary
Audio and visual equip. & services	Interview + Diary	Diary + Interview
Pets, toys, hobbies, & playground equipment	Interview + Diary	Diary
Other entertainment supplies, equipment, and services	Interview + Diary	Diary + Interview
Personal care products and services	Diary + Interview	Diary
Reading	Interview + Diary	Diary
Education	Interview + Diary	Diary
Tobacco products & smoking supplies	Interview + Diary	Diary
Miscellaneous	Interview + Diary	Diary
Cash contributions	Interview	Diary + Interview
Personal insurance and pensions		
Life and other personal insurance	Interview	Diary + Interview
Pensions and Social Security	Interview	Diary + Interview

Note: "Current Design" column from Survey Source of Data for Consumer Expenditure Survey Integrated Tables, 2009. "Interview+Diary" in the "current design" column indicates that the interview is the source of the majority of the items for this category, and that the diary is the source for the minority of the items. "Diary+Interview" indicates the converse.

As for the actual design of the CE diary, we encourage development of a diary structure that accommodates both *source* of expenditure cues as well as *type* of expenditure cues. For example, we believe that a diary that facilitates entry of information based on a respondent's preference (a single filtering question) is more likely to encourage data capture than an approach that requires adherence to a preconceived notion of categories of items. Furthermore, use of receipts follows a 'source' format rather than a 'type' format. Additionally, the landscape of where and what purchases are made have changed dramatically since the late 1960s, when the initial CE diary structure research was conducted (Sudman and Ferber 1971). For some purchases, a respondent may want to begin the recording process by answering a "what" question, with subsequent questions moving down a hierarchy until the necessary level of detail is recorded (e.g., what, fresh/frozen/bottled/canned/other, gift vs. not a gift, for males/females, age ranges). However, just as likely is a respondent who sits down to record expenditures from a "big box" shopping trip where the diary may only point to a receipt that has been scanned. Furthermore, it expands the information for 'source' currently collected in the diary of where food and drinks away from home were purchased (fast food/ full service/ vending machine/ cafeteria) to other items. We note that these are different from the Point of Purchase questions added to the April 2011 CEQ (Brattland 2011) in that the source questions would be integrated throughout the questionnaire as a memory cue, not simply an additional section to use as a benchmark for another survey.

Income and asset questions in the CE are currently obtained through the Interview. We recommend a similar approach to gathering this information. The CE data users indicate discrepancies in income and taxes relative to national benchmarks (Henderson et al. 2010). Importantly, data users indicate a mismatch between the reference period for expenditures (currently three months) and for income (currently one year). We recommend collecting from all CU's the past month's income in the diary placement interview and the equipment retrieval interview. Thus, one-month change can be estimated within CUs, as well as obtaining income from the month for which expenditures are collected. To

potentially improve reporting of income items, we recommend using similar sets of 'memory triggers' as described below for expenditures.

We also recommend collecting past 12 month income from all CUs, as in the current's design reference period, to facilitate analyses at the annual income level. Although this will suffer the same issues of not being contemporaneous with the diary data collection period as in the current design, it will match the data collection period for the panel cases at the time of the third interview (as with the fifth interview now). It will also be temporally prior to current expenditures, thus facilitating causal modeling.

Additional Inputs to CU Diary

In addition to the use of a tablet for the recording of expenditures via a "diary" we recommend supplementing the direct entry of items with electronic capture of individual items and receipts. With respect to the capture of individual items, we would recommend the integration of a barcode reader with the tablet. Although the camera capabilities of most tablets allows the device to serve as a barcode reader, the size of the tablet in most cases makes it a bit awkward for use as a barcode reader. We have explored both the use of barcode readers that are stand-alone devices (e.g., Scanfob 2002 Mobile Wireless Laser Scanner, http://serialio.com/products/scanner/mobile/Scanfob_2002.php) and various existing apps for Android and iOS devices (e.g., Barcode Scanner; RedLaser; ShopSavvy). All of these devices and apps read UPC codes and/or QR codes. The major advantage of stand-alone scanners is the ease of use and the storage capacity of the devices –for example, the Scanfob has the capacity to scan and store up to 20,000 barcoded items, batch uploading via Bluetooth to an iOS or Android-based device, with software that can be modified for specific uses via the SerialMagic API.[24] The major advantage of utilizing apps for Android and iOS devices is the potential familiarity with the apps among the population. In addition, placement of a device such as an iPod Touch with a CU for the

---

[24] We are not recommending a particular tablet, barcode reader, or scanning hardware or application. Rather the discussion is intended to offer illustrative examples of the capabilities that currently exist.

UWM
University of Wisconsin MILWAUKEE

Abt SRBI

N

purposes of scanning also provides the CU with an additional device that can be used as a portable diary by one of the CU members.

The major advantage of the use of a barcode reader is the capture of detailed information for those items with a barcode; the major disadvantage is that the barcode data do not include price information. The ability to modify the barcode data with information about costs would need to be developed.

We also recommend the integration of either a receipt reader with the tablet or a stand-alone scanner for electronic capture of receipts. For example, ProOnGo (www.proongo.com) allows users to electronically capture receipts and store the receipts via an Excel spreadsheet. The app includes a secure web portal for emailing receipts and offers users support to organize their expenses in so as to be most useful for the user. Similarly, hardware such as NeatReceipts allows for the scanning of receipts and the conversion of those receipts to spreadsheets. Although software can create expense reports from the receipts the expense reports include the total costs, not the itemized costs.

A major advantage of the use of receipt scanners is that they capture details about the item (albeit in an abbreviated format), cost of the individual items, and the details associated with discounts and coupons. Some mechanism –either electronic or human --- would need to convert scanned receipts to a listing (and associated cost) of individual items. The current CE diary tells respondents 'we need you to actually write the information *in* the diary.' Receipt scanners, however, permit flexibility in who enters the information from the receipt itself. In fact, with electronic data collection, interviewers or data entry staff could update the CU's diary with information from the receipts, even if the CU does not enter it, and not have to spend time at the CU's housing unit.

Use of both barcode readers and electronic capture of receipts has been shown to be feasible and efficient means for capturing expenditure data (Leicester and Oldfield 2009; Matulewicz et al. 2011). The integration of scanners, barcode readers, and tablets for diary data collection is one area for which we recommend methodological investigations.

CU Reporters

To minimize reliance on proxy reports, we are encouraging all adults in the CU to serve as diarists. Previous research indicates that reporting by more than one CU member results in higher estimates of expenditures for the household, especially with respect to clothing and personal care expenditures (e.g., Grootaert 1986). Previous research conducted by BLS also suggests that with respect to the number of diarists within a CU, one size may not fit all (Edgar 2006); however, to the extent possible we recommend self-reporting, with exceptions made for those CUs that strongly self-advocate for the use of a single diarist. The U.K. experience as well as the experience in other countries such as Australia (Trewin 2006) suggests that it is feasible to engage multiple respondents successfully within sampled CUs.

In addition, we recommend the use of a simple portable paper diary[25] for young consumers, those 7 to 15 years of age. The UK experience suggests that inclusion of these children's expenditures had significant impacts on the expenditure estimates for particular categories of expenditures (Horsfield 2010). These paper diaries would be collected by the interviewers and entered by BLS to maintain confidentiality of the children's purchases and to alleviate extra burden on the primary CU reporter.

Field Procedures

Sampled CUs would be recruited so as to begin serving as diarists at the start of the week following placement of the diary materials (although we note that in the U.K., diarists are to begin recording on the day following recruitment). To encourage compliance, we would recommend calls by the interviewer early in the first week of serving as a diarist as well as the provision of telephone-based help lines and online "help" support. The Australian experience has interviewers check with diarists between the second and fourth day of the first week of recording (Trewin 2006); in the U.K., interviewers are encourage to make a

---

[25] Consider allowing children ages 7 to 15 to also have an ID that allows them to enter their data via the electronic diary. Nothing precludes allowing these children and adolescents from utilizing technology for the real time capture of their expenditures, but to date, we know of no field experience with such a design.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

"checking call" after the first shopping trip, on or about day 3 of the first week (Horsfield 2010).

A critical component of the initial interview is the training of the CU members – especially the primary CU "purchaser." We envision interviewers bringing a "market basket" of goods, service bills, and receipts with them so as to demonstrate and train respondents in how to use the computer and peripheral equipment. Current practices include the use of video training, face to face individual training, group training sessions, and technical support and retraining via telephone (Trewin 2006; Link et al. 2011 Matulewicz et al. 2011). However, the literature is silent with respect to how best train respondents in the use of multiple technologies, especially when all respondents will most likely not be present for the initial interview. The lack of empirical investigations with respect to effective and efficient strategies suggest that here, too, BLS would be best served by experimentation prior to finalizing a respondent training protocol.

We recommend a flexible interface with the CU during the diary month. For some CUs, we believe that multiple face to face contacts to encourage participation and to provide further training and answering questions will be beneficial. Other CUs may require little assistance or encouragement during the diary month. An advantage of a web-based data collection system is that compliance with the task can be monitored and interventions and encouragement based on the level of activity, the compliance across CU members, and the demographic composition of the CU. Recent research has shown that feedback and probing can be used on open-ended questions in web surveys to result in an improvement in data quality (Oudejans and Christian 2011). Thus, we recommend a multi-pronged strategy, with real-time feedback built into the web or tablet instrument and follow-ups for CUs who have less experience with technology, who show lower compliance, or who request additional assistance.

The diary period would conclude with a face to face interview shortly after the end of the month-long panel. The interview would serve multiple purposes: (1) to collect the data collection equipment; (2) to administer retrospective questions on topics that are prone to underreporting in the diary, for example,

monthly payments of mortgages, utilities, or automatic payments; and (3) to review the materials for completeness, with options to administer retrospective question modules as necessary and to disentangle scanned receipts into individual, codeable items. Figure 2 summarizes the field procedure protocol as outlined here.

```
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Interviewers     │      │ Initial          │      │ Diary placement  │
│ contact CU and   │ ──►  │ interview: CU-   │ ──►  │ and initial      │
│ establish        │      │ level            │      │ completion:      │
│ eligibility      │      │ characteristics +│      │ Week 1           │
│                  │      │ training         │      │                  │
└──────────────────┘      └──────────────────┘      └──────────────────┘
                                                              │
                                                              ▼
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Week 4: Diary    │      │ Week 3: Diary    │      │ Week 2: Diary    │
│ completion;      │ ◄──  │ completion;      │ ◄──  │ completion;      │
│ Follow-up calls  │      │ Follow-up calls  │      │ Follow-up calls  │
└──────────────────┘      └──────────────────┘      └──────────────────┘
         │
         ▼
┌──────────────────┐
│ Day 30: I'ers    │
│ pick up          │
│ equipment,       │
│ review receipts, │
│ final            │
│ questionnaire    │
└──────────────────┘
```
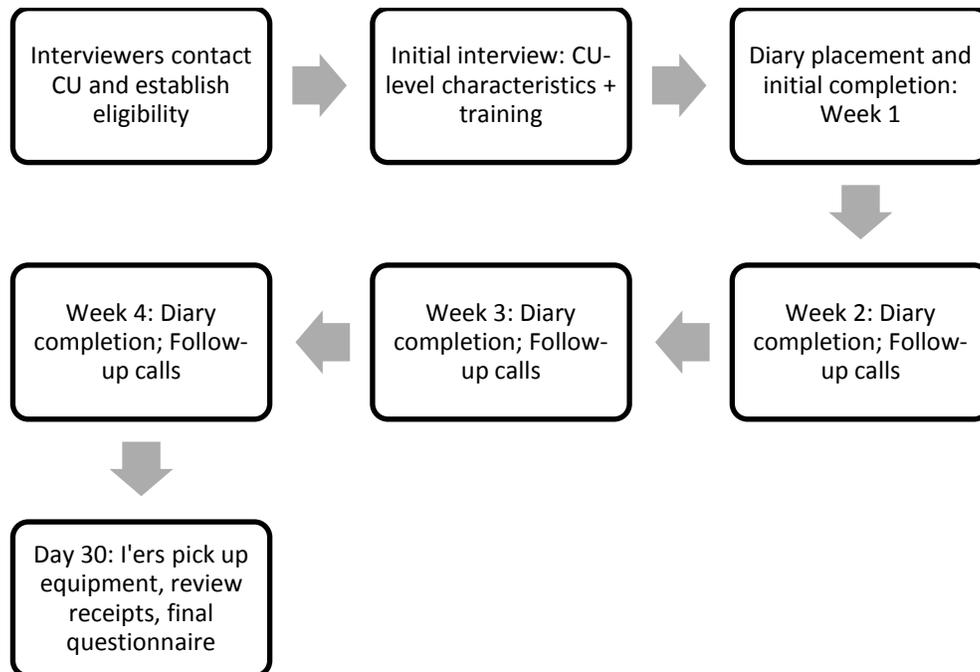
Figure 2: Diary Questionnaire Implementation Procedures

Memory Triggers

As a means for minimizing the reliance on retrospective recall, our design recommends the use of a "portable" diary for all adults ages 16 and older for which the form of the diary is determined by the individual CU member. A recent study found that 83% of U.S. adults own some type of cellphone and that 73% of cellphone owners use text messaging (Pew 2010). There is a growing body of empirical literature the demonstrates the feasibility of using cellphones, smartphones, and PDAs for "real time" capture of behavioral data – via pictures,

audio recordings to oneself, and text or voice messages (e.g., Palen and Salzman 2002; Hosoe 2005; Marceau et al. 2007; Arens et al. 2008; Sun et al. 2010; Palmblad and Tiplady 2004).   These "portable" diaries are intended to capture bits of data to serve as memory cues for the respondent to use when entering data into the diary. With respect to these memory triggers, we recommend a broad range of modes of data capture be offered – small paper diaries, interactive voice recognition "drop boxes" (where respondents can dial in and record information about purchases), smartphone apps for voice or picture "diaries" of locations or purchases. As with the use of financial tracking software, here we advocate for a flexible design that reflects the technological facileness and comfort level of individual CU members and which builds on whatever means they may already be using to track expenses.[26]

Electronic receipts for internet purchases, for brick and mortar purchases, and for services rendered, offer a convenient means of data capture. How best to take advantage of these receipts? Once again, our recommendation is to minimize the need for the CU respondents to have to re-enter data that are already in an electronic form. As with scanned receipts, the most efficient approach (from the perspective of the respondent) may be to have the capability to upload such receipts directly to the diary data record. Another means by which to capture these electronic receipts may be to have a secure file location where each CU member can forward his or her receipts, tag them, and then reference that tag in their diary entry for the day. Although this puts additional work in the hands of the interviewer or a coder, it reduces the burden for the respondent.

For those CUs that already monitor expenditures, here we advocate for a flexible design. The goal in developing a flexible design is to utilize the data that are already captured by the CU (hence reducing redundancy in data collection efforts) while capturing the full range of expenditures of interest to BLS. The 2011 Consumer Financial Literacy Survey (Harris Interactive Inc. 2011) found

---

[26]In addition to using electronic records as memory devices, we believe it is feasible for many of these electronic records to be imported directly to a diary-based record.

that 43% of adults in the U.S. have a budget and keep close track of their spending, although no data were collected on the use of software to monitor expenses. Data from a YouGov survey (self-selected sample) indicate that 40% of those with mobile phones, smartphones, or tablets use mobile banking services (Antenna 2011). In its review of financial software, Westat (2011) noted that financial tracking software contain several features that are relevant to the CE: registers that display transactions and amounts, interfaces that facilitate downloading data from financial institutions, and the ability to split and label transactions into identifiable, individual items (e.g., the Target receipt for personal care items and clothes can be decomposed). Financial tracking software is also available for use with smartphones, allowing real time entry for those respondents attempting to monitor their expenditures. For example, mint.com provides a point-of-purchase application that allows users to record at the time of the purchase and directly input those data to the financial management software. To the extent that some households are, in essence, already capturing CE data, every effort should be made to translate these data to a format usable to the CE. How best to do this and meet the needs of the CE is one area for which we suggest further investigation (see Section IV).

To improve the reporting of income sources, we recommend using multiple record sources to collect information about income as well. Since the information will be collected at the end of the diary data collection period, the CU has the opportunity to organize records and record the information along the way. In fact, the CU can be told to simply keep records for anything related to money coming into the household or leaving the household, thus unifying the two concepts with one task. For example, a waitress could record tips in her pocket diary at the end of her shift, a professional could forward an electronic pay statement to their electronic study email account, and a small business owner could use output from their business financial software to output relevant information for the month. In this way, the task at hand for income is not meaningfully different from expenditures, although it may be more sensitive. Because of this, two approaches could be used. First, the interviewer could

directly ask about a variety of income sources and taxes paid as currently done in the survey. Before they leave, the CU could also complete a final self-administered interview where they are again asked about income in an attempt to overcome the sensitivity of the topic (e.g., Tourangeau and Smith 1996).

Engaging the Respondent

The request made of selected consumer units is not trivial. We recommend that the BLS investigate the use of incentives, feedback mechanisms, or both, as part of the field protocol to be used for the CE. The literature demonstrates that incentives are effective and efficient (e.g., Singer 2002) while the precedence for the use of incentives exists both for Federal expenditure surveys in other countries (Horsfield 2010) and for high burden U.S. Federal surveys   The U.K. pays each adult respondent 10£ and children who maintain diaries 5£. Nielsen Homescan panel members are not offered incentives but rather earn "points" that can be redeemed for merchandise. The National Health and Nutrition Examination Survey (NHANES) offers an incentive for participating in the medical examination portion of the study, once again, offering the incentive at the person level, not the household level, so as to improve participation among all household members.

How else can the CE work with respondents and engage them in the process of providing this valuable data? One means of engagement may be to provide respondents with feedback concerning expenditures. Findings from the 2011 Consumer Financial Literacy Survey indicate that 42% of adults give themselves a grade of C or lower with respect to knowledge about personal finance and 76% feel they could benefit from financial advice and answers to everyday financial questions (Harris Interactive 2011). NHANES provides respondents who participate in the medical examination portion of the survey with a report of medical findings, thus providing some feedback of potential benefit to the respondent. Feedback to the CU could include a summary of their expenditures by major item, changes in their expenditures over the course of the month, and perhaps comparison to other households of similar size, region, and

household composition. Although we are not advocating for BLS to provide financial advice, the responses to the Consumer Financial Literacy Survey suggest that there may be opportunities for information exchange between the data collector and the data provider. Here, we do not have a concrete recommendation but rather encourage BLS to engage with CU respondents to learn what means of engagement or feedback information would be of value to those CUs.

C. Alternative Data Sources

Regardless of the extent to which technology is incorporated in the data collection process, the CE is and will remain a burdensome survey in which to participate. How can BLS effectively use alternative data sources to either reduce respondent burden or improve estimation? We address the uses of alternative data sources with respect to issues related to nonresponse in Section V. In this section, we discuss various uses of alternative survey and administrative data as a means of reducing respondent burden or for purposes of assessment of measurement error.

Alternative survey and administrative data can be used for two broadly defined purposes: (1) at an aggregate level to serve as benchmarks (a 'top down' approach; Heimovitz 2003) or (2) at the micro-level for purposes of supplementing or replacing CE data (a 'bottom up' approach; Heimovitz 2003). With respect to micro-level use, further consideration must be given to direct linking of records or data (that is, at the person or household/CU level) or synthetic linkage, whereby CE data are augmented with data from statistically "similar" consumer units or individuals.

The use of administrative records to augment survey data is a growing area within survey methodology (e.g., Statistics Netherlands 2000). Studies indicate significant issues in direct linkage of survey and administrative records, governmental and not, ranging from difficulties obtaining consent to do the

linkage to inaccuracies in the linkage itself (Raghnathan and Van Hoewyk no year; Heimowitz 2003; Jenkins et al. 2004; Sakshaug 2011).

Except for those cases in which direct linkage at the micro-level is possible (e.g., direct acquisition of a respondent's bank account or credit card records, tax return, or records of transfer benefits), a CE design that attempts to reduce respondent burden through the use of alternative sources of expenditure data moves BLS away from a single, integrated data source that is desired by both CPI and outside data users. It also potentially decreases timeliness.

Given the current state of the field (both with respect to quality of data linkages and quality of administrative record sources), our recommendation at this time is to consider use of alterative data sources not directly but rather as benchmark data to gain greater understanding of measurement error, nonresponse error, or both.[27]

At least three Federal surveys provide detailed information that could be utilized by BLS as either alternative sources of CE-type data or as benchmarks. The three surveys are the Medical Expenditure Panel Survey (MEPS), the National Household Food Acquisition and Purchase Survey (FOODAPS), and the Residential Energy Consumption Survey (RECS). Each of these surveys collects detailed information at the dwelling unit or household unit level with respect to expenditures for medical expenses, food, and energy use, respectfully.

The Medical Expenditure Panel Survey is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States. MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of health insurance held by and available to U.S. workers. MEPS currently consists of two major components: the Household Component and the Insurance Component. During the household interviews, MEPS collects detailed information for each person in the household on the following:

---

[27] We of course recognize that none of the data sources discussed in this section are a true gold standard. As is true for all administrative and survey data, each is subject to various sources of error that must be evaluated before its use or integration with CE.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE   Abt SRBI   N

demographic characteristics, health conditions, health status, use of medical services, charges and source of payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The Insurance Component collects data from a sample of private and public sector employers on the health insurance plans they offer their employees. The survey is also known as the Health Insurance Cost Study. The collected data include the number and types of private insurance plans offered (if any), premiums, contributions by employers and employees, eligibility requirements, benefits associated with these plans, and employer characteristics.

The National Household Food Acquisition and Purchase Survey (FoodAPS) will provide detailed data about household food choices. FoodAPS will be a nationally representative survey of household food purchases and acquisitions. Detailed information will be collected about foods purchased for consumption at home and away from home as well as foods acquired through food and nutrition assistance programs (both public and private). Of primary interest to the CE program is the focus of the survey on the quantities, prices, and expenditures for all at-home and away-from-home foods and beverages purchased and acquired from all sources. The sample size, however, is anticipated to be relatively small (and therefore may not meet the needs of the CE program) collecting information from up to 3,500 low-income and 1,500 higher income households and, to our knowledge, will be a one-time survey.

The Residential Energy Consumption Survey (RECS) (conducted by the Energy Information Agency) collects data on energy use from a nationally representative sample of housing units. Information is collected on energy characteristics of the housing unit, usage patterns, and household demographics. This information is combined with data from energy suppliers to these homes to estimate energy costs and usage for heating, cooling, appliances and other end uses. First conducted in 1978, the thirteenth RECS was conducted in 2009. The 2009 survey collected data from 12,083 households in housing units statistically selected to represent the 113.6 million housing units that are occupied as a primary residence.

Direct use of any of these surveys would involve surmounting issues of definitional compatibility and reference period correspondence; consideration of use of these records at the micro-level involves further issues concerning data sharing across Federal agencies. As such, we consider these as sources for aggregate benchmarks rather than sources for micro-level data.

Beyond data available from Federal surveys which focus directly on some aspect of household expenditures, there are numerous data sources which could be used to supplement or assess CE data. Customer loyalty programs offer one such data source; several studies have incorporated the use of this type of data to examine measurement error in consumer expenditure surveys (e.g., Einav et al. 2008). Some market segments offer data sources that aggregate across distributors, to produce national level estimates of utilization and expenditures, the best example of which is IMS Health, which aggregates prescription medicine expenditures across more than 50,000 pharmacies (IMS 2011). Various corporations collect consumer expenditure data – the best examples of which are the Nielsen HomeScan Data, now the National Consumer Panel (http://www.ncppanel.com/) and the Experian Simmons National Consumer Study (Experian 2011). Other firms provide aggregate data and analysis of particular market segments (e.g., CardData, Acxiom). These studies collect information on types and brands of various consumer goods; Nielsen Homescan also collects information on prices. The samples on which these data are based vary widely -- from self-selected samples to national probability based sample designs -- as do the data collection methodologies (paper diaries to high tech barcode scanning of all items brought into the household).

To what extent would data from commercial vendors of consumer expenditures be of use to BLS? After reviewing data quality evaluations of these various data sources (e.g., Einav, Leibtag and Nevo 2010), we do not recommend any one of these data sources as a replacement or supplement to the CE. Rather they offer data that may, for particular expenditure categories offer further insight into consumption patterns not possible with the CE data and thus be suggestive of information to include in nonresponse adjustments and

monitoring for data quality issues. For example, comparisons of the Homescan file with grocery store records show that location, type, and quantity of purchase have relatively high correspondence with store records, but the price does not. The mismatch between reported and store data appears to be due both to the consumer and to imputation errors. Households who reported the number of trips made to retail outlets accurately were smaller households, those with fewer children, married, and those in which the female was not fully employed and had lower levels of education, on average (Einav, Leibtag and Nevo 2010).

D. Risks

The most significant risk associated with our recommended design relates to the extension of the diary period, from a two week period to a one month period. At present, there are no contemporary empirical data (of which we are aware) that addresses (experimentally) the impact of an extension of the participation period for a national consumer expenditure survey. There are clearly examples of panelists in private sector consumer studies who participate for periods far in excess of a month; whether that level of compliance can be obtained in a Federal data collection effort is key to the success of the recommended design. The importance of this feature – coupled with the lack of experimental or even observational data – is why we strongly recommend the methodological investigations concerning compliance outlined in Section IV.

There are also significant risks associated with the implementation of our design at the operational level. Is it feasible to place tablet technology in households for a month? What is the failure rate for the various pieces of electronic equipment and the communication among these devices? How durable is the equipment? What security measures need to be put in place both with respect to the physical equipment as well as the electronic capture of information? Much of the research that places technology with participants for a longer period of time is of small samples or highly engaged populations (e.g., Cook et al. 2011; Lai et al. 2010).

The third major area of risk associated with our design recommendation concerns the data user. We are cognizant that some users require quarterly and annual estimates of total spending by a consumer unit (e.g., Department of Defense; see Henderson et al. 2010). Balancing the needs of various users is probably the most difficult task facing BLS. We have chosen to give highest (and to some extent) sole priority to the improvement of the data for the CPI. One could consider a number of permutations to our recommended design that would facilitate meeting these non-CPI users' needs, but they all come at significant costs. For example, shifting the recall period from a three month recall to a one month recall, coupled with extensive encouragement of retention of receipts and interviewing CUs every month for three months, then nine months off, followed by a second three-month panel period would provide the Department of Defense with the quarterly data they need and would provide users with year to year assessments of CU-level change in expenditures. The one month recall could be validated against the one-month diary period, under the assumption that the diary provides a more accurate picture of expenditures. We considered multiple design options so as to meet the calendar year data; none of the options we considered provided BLS with a cost-effective means by which to capture these data while reducing measurement error associated with retrospective recall.

IV.   Methodological Investigations to Support Design Decision

The recommended design outlined in Section III represents a radical departure from the current CE survey. We have attempted to support our recommendations with citations from the empirical and theoretical literature, as well as our understanding of best practices currently in use. However, much of the empirical literature on which we base our recommendations is dated and often times references survey tasks quite dissimilar to the task facing respondents to the consumer expenditure survey. In addition, some of our recommendations are based on cutting edge practices for which there is not an experimental or empirical literature to reference. We recognize that the original request for proposals indicated that recommendations for methodological investigations were to be held to a minimum. We have attempted to honor that requirement, but we feel we would be remiss as methodologists if we failed to point out where our recommendations are weakest with respect to current empirical knowledge.

Our recommended design outlines the need for further investigation and experimentation with respect to the following key features of the redesigned Consumer Expenditure Survey:

- Length of the diary period and effects of the use of technology;
- The design of respondent training protocols; and
- The extent and level of incentives.

Apart from these key design features, we recommend smaller studies to examine the following factors:

- The use of multiple formats of memory trigger diaries;
- The level and frequency of interviewer-respondent interaction over the diary month;
- The integration and use of respondents' financial management and tracking systems so as to reduce respondent burden; and
- The incorporation of feedback mechanisms for responding consumer units.

To address the three key design features, we offer below investigations we believe would be informative to BLS as it moves toward a redesigned expenditure survey.   We have attempted to be cost efficient by designing the experiments described below so as to also serve as feasibility studies.

Experiment #1: Length of Diary and Effect of Technology Experiments

The design recommended above calls for a diary recording time that is significantly longer and significantly different than the current CE diary instrument. Given the costs involved with training the consumer unit to use the tablet as a data collection instrument and the desire of data users to have more information on each household, we believe that this design will be a reasonable balance of survey errors and survey costs. However, early research (e.g., Sudman and Ferber 1971; Sudman and Bradburn 1973) suggested that a four week diary may be too long for consumer units to reliably comply with the request to record information in the diary. Because this experiment will be used to establish the length of the diary, it is important to disentangle the effect of the diary length from the effect of technology. That is, does the electronic diary proposed here do better or worse than a comparable paper diary of the same length. We believe that these two factors should be fully crossed to examine the main effects and interaction effects.

The following two experimental factors are:

1. The length of the diary data collection period: 1 week, 2 weeks, 4 weeks, or 6 weeks
2. Electronic diary vs. paper diary

Factor A: Length of the diary field period

In this experimental factor, the consumer unit will be asked to record the diary for one week, two consecutive weeks, four consecutive weeks or six consecutive weeks. At the beginning of the interview, consumer units will be instructed that they will need to record expenses in the tablet diary. The tasks will be identical for each of the weeks, but will vary in total length.   We anticipate

that compliance will fall as the length of the diary time increases. To understand whether four weeks is the maximum amount of time, this factor increases the diary period to six weeks for a random sample of households.

In particular, the first experimental feature would be as follows:

Group 1: Interviewers conduct the initial interview as outlined Section III above. The tablet with the web diary or the current paper diary[28] is placed in the CU. The CU is instructed to record all expenditures for a total of one week. Interviewers make follow-up calls as appropriate to encourage participation. The interviewer returns on day 7 to collect the equipment/diaries and conduct the final interview.

Group 2: Interviewers conduct the initial interview as outlined in Section III above. The tablet with the web diary or the current paper diary is placed in the CU. The CU is instructed to record all expenditures for a total of two weeks. Interviewers make follow-up calls as appropriate to encourage participation. The interviewer returns on day 14 to collect the equipment/diaries and conduct the final interview.

Group 3: Interviewers conduct the initial interview as outlined in Section III above. The tablet with the web diary or the current paper diary is placed in the CU. The CU is instructed to record all expenditures for a total of two weeks. Interviewers make follow-up calls as appropriate to encourage participation. The interviewer returns on day 28 to collect the equipment/diaries and conduct the final interview.

Group 4: Interviewers conduct the initial interview as outlined in Section III above. The tablet with the web diary or the current paper diary is placed in the CU. The CU is instructed to record all expenditures for a total of two weeks. Interviewers make follow-up calls as appropriate to encourage participation. The interviewer returns on day 42 to collect the equipment/diaries and conduct the final interview.

---

[28] We recommend the use of the current paper diary to reduce costs associated with developing both a computer/web-based instrument and a revised paper document.

UWM
UNIVERSITY of WISCONSIN MILWAUKEE

Abt SRBI

N

Factor B: Mode of diary

Group 1: Tablet diary: Respondents are given the tablet diary with the web-
                based questionnaire, scanner and bar code reader described above.

Group 2: Paper diary: Respondents are given the current paper diary.

Practical Insights into the New Design: A major unknown is whether CUs will comply with the task of keeping expenditure and income records in an electronic diary for four weeks. It is also unknown how a shift from a paper and pencil diary to an electronic diary system will affect participation and data quality. That is, this task is substantially different from the current design in that it is both longer and electronic. As such, this experiment permits disentangling the effect of length from the effect of technology through experimental design.

    As with the feasibility laboratory study described below, outcomes to be examined include durability of the equipment, rate of failure of the devices during the diary period and whether the failure rate changes as the diary reporting period increases, and safety/security of leaving equipment of this value in CU members' homes. Other study-related outcomes could include diary cooperation rates per week, item missing rates, number of goods reported, number of receipts scanned; number of items with bar code readers used; number of e-receipts sent to email; number of individual diaries used; and total dollar amount reported in each expenditure category, following data quality indicators recommended by Gonzalez, et al. (2009).

Estimated Financial and Time Requirements:

        Early research (Sudman and Ferber 1971) indicated that cumulative response rates fell by roughly 10 percentage points for each additional week of keeping a diary. Simple random samples of about size n=400 for each main effect would be sufficient to detect a difference in response rates of 60% and 50% at an alpha<.05 level with power at least 0.8. If, however, the difference is only 5 percentage points (e.g., 50% and 55%), then samples of size n=1565 are needed in each group for the main effects. With the latter in mind, and assuming that pairwise comparisons of the main effects of each factor are of interest rather

than simply tests of overall effects, this would recommend a simple random sample size of roughly 6300 to detect differences between these eight groups of 5 percentage points.

Assume, based on the power calculations above, a sample of size 6300 CUs is selected and evenly divided between the eight fully-crossed treatments. This yields n=787 or n=788cases in each cell. We assume that initial compliance rates differ somewhat between the four treatments for the first experimental factor of length (Sudman and Ferber 1971), but have no clear expectations for the second experimental factor of technology on nonresponse. For purposes of planning, we assume that the initial compliance for Group 1 is 80%, for Group 2 is 75%, for Group 3 is 70% and for Group 4 is also 70%. This would yield 1260 initial respondents in Group 1, 1182 in Group 2, and 1102 to 1104 each in Groups 3 and 4, for a total of 4648 initial respondents. The size of the initial pool is important for cost reasons; we anticipate that the final cumulative compliance rates for Groups 2, 3, and 4 will fall by 5 to 10 percentage points per week. See Table 2 for a summary of this design.

Table 2: Expected Sample Size for the Eight Length by Mode Experimental Cells

	1 week	2 weeks	4 weeks	6 weeks	Total
Electronic	787 cases; 630 initial Rs	787 cases; 591 initial Rs	788 cases; 552 initial Rs	788 cases; 551 initial Rs	3150 cases; 2324 initial Rs
Paper	788 cases; 630 initial Rs	788 cases; 591 initial Rs	787 cases; 552 initial Rs	787 cases; 551 initial Rs	3150 cases; 2324 initial Rs
Total	1585 cases; 1260 initial Rs	1585 cases; 1182 initial Rs	1585 cases; 1104 initial Rs	1585 cases; 1102 initial Rs	6300 cases; 4648 initial Rs

What about power to detect differences in reporting among the respondents? It is difficult to know what is an appropriate effect size to use for power calculations in this experiment. Edgar (2010) found that respondents in the CE report an average number of 29.4 items with a standard deviation 17.5 items. Assuming that the within-cell standard deviation for the number of items is equal to 17.5 items (Edgar 2010), and that there are 581 respondents per cell (4648/8;

we acknowledge that the actual design will be unbalanced due to differential nonresponse as outlined in Table 2), then we could detect an interaction effect between the length of reporting and mode of data collection of 4 items with power of 0.81. [29]

Cost in this experimental factor is driven by the cost of acquiring the tablets, barcode readers, and scanning equipment and programming the surveys. Our current estimates of the cost of this equipment (see Section VI below) for retail prices, not including bulk or government discounts, is $900 to $1500 for all of the pieces of equipment per CU. With 4648 total responding CUs, this yields a total equipment cost for this experiment to be between $4 million and $7 million should all equipment be used simultaneously. Of course, this investment in equipment is not for naught, as the equipment, should it be feasible, will be rolled into the main study.

As in the approach of the first feasibility study described below and to reduce overall initial investment costs, the sample could be released in replicates for four consecutive months. This would reduce the total number of expected respondents to be 4648/4 = 1162 respondents, with a corresponding reduction in the number of tablets, scanners and barcode readers to purchase. In this design, then there would be approximately 145 persons per experimental cell per month. This approach would reduce the cost of equipment for the study to roughly $1 million to $1.7 million.

Assuming a cost of $325 per completed interview, the total field interviewing cost would be around $1.5 million.

Are these costs too large for a field test? We harken to criticism of attempts to integrate technology into the 2010 and for previous 'pilot' designs that confound methodological factors and fail to do power calculations Census (see, e.g., Brown, Cohen, Cork and Citro 2010, pp. 71-73). Our recommendations here follow those presented in a previous CNSTAT panel's report on the 2010 Census, in which we have identified function rather than actual form. We have no

---

[29] We note that other estimands and other analyses will have different power calculations. We focus on the total number of items as one important data quality metric that we expect to be affected by the design, using a 'more is better' heuristic.

UWM
University of Wisconsin MILWAUKEE

Abt SRBI

N

way of knowing what the future will hold in terms of tablet computers or if they will morph into yet another form of technology. However, a set of technology that can do these tasks is likely to save costs in the long-run, especially related to post-survey processing and data entry.

We have also attempted, using the best data we have available, to estimate sample sizes for this experiment with reasonable amount of statistical power that permits important scientific questions to be answered. The reason that these two factors are experimentally crossed is that the length of the diary and use of technology decision are the two ways in which this design differs from the current CE. Any study that examines one of them without the other fails to unconfound these factors from the current design.

Experiment #2: Technology Training

This study will permit an initial laboratory-based feasibility evaluation of the tablet computers as well as methods for training respondents. How to train sampled consumer units to use the tablet and scanner for recording their information is also an open question. The organizations we contacted and literature we reviewed either used in person training (Nielsen in South Africa; RTI; FOODAPPS experience) or DVDs or videos (Nielsen in the US;). The CE is a large, national US survey with, we hypothesize, a more heterogeneous population than examined in these previous studies (e.g., larger sample of elderly). We anticipate that some consumer units will need extensive training, while others will catch on right away. To test this, we propose a laboratory-based experimental test of the following three training methods:

Group 1: Video: In this method, a brief (15 minute) video training session will be preloaded on the tablet computers. Each consumer unit will be instructed on the basics of keeping records, scanning receipts, reading bar codes, emailing e-receipts, calling the in-bound IVR line, and entering information into the diary. Consumer units will be able to access the video instructions, along with a text document containing the video's script, at any time. The video will demonstrate how to enter a "market basket" of goods, service

bills, and receipts using the tablet computer and peripheral equipment. Interviewers will be present while the video is playing, but will not actually instruct the respondents on how to use the tablet.

Group 2: In person, basic: In this method, interviewers will provide each consumer unit with a brief (15 minute) training session on how to use the tablet computer and other survey tasks. As in the video, each consumer unit will be instructed on the basics of keeping records, scanning receipts, reading bar codes, emailing e-receipts, calling the in-bound IVR line, and entering information into the diary. Consumer units will be given a text document containing the basic instructions. The interviewer will demonstrate how to enter a "market basket" of goods, service bills, and receipts in how the tablet computer and peripheral equipment, but will not engage in extensive one-on-one time with the consumer unit using the tablet.

Group 3: In person, extended: In this method, interviewers will provide each consumer unit with an extended (30-45 minute) training session on how to use the tablet computer and other survey tasks. As in the video and brief training, each consumer unit will be instructed on the basics of keeping records, scanning receipts, reading bar codes, emailing e-receipts, calling the in-bound IVR line, and entering information into the diary. Consumer units will be given a text document containing the basic instructions. Interviewers will demonstrate how to use the tablet. The interviewer will demonstrate how to enter a "market basket" of goods, service bills, and receipts in how the tablet computer and peripheral equipment. They will also engage in practice sessions with the respondent, showing them how to scan a practice receipt, asking them to call in to the inbound IVR line to practice memory cues, and walking through entering information in a practice diary from three receipts from different locations.

Laboratory evaluation: The laboratory evaluation consists of three steps. First, a convenience sample of respondents will be recruited to come into the laboratory.

These laboratory respondents are randomly assigned to one of the three training treatments above. The protocol in the laboratory will be as follows:

Step 1: Administer the first questionnaire for the field data collection to 'set the stage' and obtain the same information as would be collected in the full protocol as described above.

Step 2: Administer the assigned training method, described above, to introduce the respondent to the technology.

Step 3: Immediately following the training, provide respondents with a set of items, records and receipts to enter into the diary. These items, records and receipts should reflect approximately the number of items and types of items purchased during a normal week of diary keeping, using the current CE diary and interview as a guide to identify relevant items. For example, this could include a few sacks of groceries, bags from a warehouse or multi-category retail 'big box' store including purchases that fall into a variety of categories (such as Costco, Walmart, Sam's Club and/or Target), a receipt for a haircut, utility bills, mortgage or rent bills, gasoline receipts, paychecks, and bank statements that reflect automatic transfer payments. The respondent will be instructed to scan items using the barcode scanner, scan receipts, and complete the appropriate sections of the diary.   Since the items and receipts will be known, this provides a direct measure of accuracy of entering information across the three training groups. However, it lacks external validity. As such, we include Step 4.

Step 4: Each lab respondent will provide their home address. The lab interviewer will travel to the lab respondent's house to deliver the equipment. The lab interviewer will identify whether or not the respondent has a high-speed internet connection that can work with the equipment; otherwise, the interviewer will provide the household with a tablet that will permit wireless connection to the internet without a phone line. Each laboratory respondent will be asked to keep an expenditure diary for a total of two weeks. The laboratory interviewer will make follow-up calls

and visits, as needed, to troubleshoot problems. At the end of the two weeks, the interviewers will return to the lab respondent's house to pick up the equipment.

We suggest evaluating the following outcomes during the field period: Diary cooperation rates per week, item missing rates, number of goods reported, number of receipts scanned; number of items with bar code readers used; number of e-receipts sent to email; number of individual diaries used; and total dollar amount reported in various expenditure categories.

Practical Insights into the Proposed Design: One open question for moving to a highly technological study is how to train respondents appropriately to use these technologies. This study provides both a feasibility study as well as a training study – can the equipment be used, will it malfunction, and how best to train people to use the equipment.   The three training methods range from least costly, but perhaps least effective, to most costly in terms of interviewer time and perhaps also respondent burden, but we hypothesize, probably the most effective.

The two components of this study – the in-lab entering of simulated purchases and the at-home entering of actual purchases – allow different questions to be answered. In the lab, we have a direct measure of accuracy since we get to create 'truth.' We also minimize the risk of computer failure in the laboratory. By accompanying it with a follow-up in-home diary component, we get an initial test of how feasible it is to carry out this study. The two week time frame suggested here for the field feasibility test is to permit a long enough time for fatigue to set in, but also permit greater rotation of the equipment across lab respondents.

We also can evaluate the risk in placing such costly equipment at respondent's own homes. Recruiting the lab respondents in three waves and reusing the equipment (described below), although it lengthens the time for this feasibility experiment, also permits an evaluation of difficulty in cleaning any

information potentially stored locally from the machines (a necessary part of the actual study), logistics in setting up the equipment, and the durability of the equipment, along with other potential problems.

Estimated Financial and Time Requirements: It is especially difficult to know what the effect size for respondent training should be to use for power calculations in this experiment. Again using Edgar's(2010) findings that the average number of items is 29.4 with a standard deviation 17.5 items as a baseline, assuming that this laboratory experiment and feasibility study is smaller scale than a full field test, and assuming that pairwise comparisons between the three training groups are of primary interest, we calculated the power to detect a difference between two of the groups in the total number of expenditures. These results are reported in Table 2. For example, if each of the cells has 25 respondents for a total of 75 respondents across the three groups, then there is at least 80% power to detect a difference between two of the groups if the number of items they report differs by at least 15. If the number of respondents in each cell doubles, then the minimum number of items differing between the groups that can be detected falls to 10 items. Other possibilities for sample sizes are outlined in Table 3.

For our cost estimates and for lack of any other information, we will assume a minimum difference of at least 10 items reported between two of the training groups will result. With at least 50 people in each cell, we can detect this difference with a power of 0.81. Thus, there will be a total of n=150 lab respondents. The majority of the cost for this study will come in the programming time and the equipment costs.

Table 3: Power to detect a difference in means between independent groups, assuming both groups have the same standard deviation = 17.5 and are equal sized

True difference in means	25 each / 75 total	50 each / 150 total	75 each / 225 total	100 each / 300 total	200 each / 600 total
1	0.05	0.06	0.06	0.07	0.09
2	0.07	0.09	0.11	0.13	0.21
3	0.09	0.14	0.18	0.23	0.40
4	0.12	0.20	0.28	0.36	0.63
5	0.17	0.29	0.41	0.52	**0.81**
6	0.22	0.40	0.55	0.67	**0.93**
7	0.28	0.51	0.68	**0.80**	0.98
8	0.35	0.62	0.79	**0.90**	**1.00**
9	0.43	0.72	**0.88**	0.95	**1.00**
10	0.51	**0.81**	0.94	0.98	**1.00**
11	0.59	**0.88**	0.97	0.99	**1.00**
12	0.66	**0.92**	0.99	**1.00**	**1.00**
13	0.73	**0.96**	0.99	**1.00**	**1.00**
14	0.79	**0.98**	**1.00**	**1.00**	**1.00**
15	**0.84**	**0.99**	**1.00**	**1.00**	**1.00**
16	**0.89**	**0.99**	**1.00**	**1.00**	**1.00**
17	**0.92**	**1.00**	**1.00**	**1.00**	**1.00**
18	**0.95**	**1.00**	**1.00**	**1.00**	**1.00**
19	**0.96**	**1.00**	**1.00**	**1.00**	**1.00**
20	**0.98**	**1.00**	**1.00**	**1.00**	**1.00**

Note: Power calculations from: Lenth, R. V. (2006-9). Java Applets for Power and Sample Size [Computer software]. Retrieved October, 2011 from http://www.stat.uiowa.edu/~rlenth/Power

To keep equipment costs under control for this initial feasibility test, we recommend recruiting respondents in three waves across three months, with n=50 respondents in each wave, equally spread out across the three training groups (e.g., n=16 in Group 1, n=17 in Group 2 and n=17 in Group 3; see Table 4). Each wave will be a new independent 'cross-section' of volunteer respondents, recruited through normal volunteer methods such as Craig's List, newspaper ads, word-of-mouth and so on. We recommend using age and/or education as blocking variables for this study to ensure variability in respondent characteristics. With this approach, then only 50 sets of equipment need to be purchased, with an equipment cost ranging from $27,000 to $45,000.

UWM
UNIVERSITY of WISCONSIN **MILWAUKEE**

Abt SRBI

N

Table 4: Training and Feasibility Study Timeline

Week	Month 1	Month 2	Month 3
1	Recruit n=50 lab participants Group 1: 16 Group 2: 17 Group 3: 17 Start Lab + Diary	Recruit n=50 lab participants Group 1: 17 Group 2: 17 Group 3: 16 Start Lab + Diary	Recruit n=50 lab participants Group 1: 17 Group 2: 16 Group 3: 17 Start Lab + Diary
2	Lab + Diary	Lab + Diary	Lab + Diary
3	Complete Diary; pick up equipment	Complete Diary; pick up equipment	Complete Diary; pick up equipment
4	Ready equipment for next group	Ready equipment for next group	End of study

Assume that a lab interview and training takes approximately 1 hour of an interviewers' hands-on time, plus approximately 30 minutes for set up and clean up (15 minutes each). This would result in a total of 225 hours of lab interviewer's time. If we assume that lab interviewers are paid $25 per hour, then the lab interviewer salary cost will be $5625 + fringe benefits. Further assume that some nominal incentive is necessary to recruit respondents into the lab; if the incentive is $25 per laboratory respondent, then this would take $25*150 respondents = $1250 in incentive funds.

Assume that the lab respondents are recruited from the same general metro area. Costs for interviewers driving to the respondent's house to drop off, set up and pick up the equipment involve mileage and time. Assume that respondents live, on average, 30 miles from the laboratory, and that reimbursement is the current 51 cents per mile. Over all of the respondents, each trip requires 30 miles * 150 respondents = 4500 miles or $2295. Since there are four such trips, the total mileage costs are $9180. Assume that traffic moves at 30 miles per hour; each trip then takes 1 hour, for a total of 4*150=450 hours, or a total amount of $11250 in interviewer time. Further assume that it takes about 30 minutes to set up and test the equipment at the respondent's house. This would be, at $25/hour, an additional $1875 in interviewer time.

Excluding equipment costs, this totals to roughly $29200. We will round this up to $30,000. Including equipment costs, the total cost of the experiment

will be between $57,000 and $75,000. Note, however, that this does not include the initial costs or time of programming the instruments.

We place this experiment as the second recommendation, even though we estimate the costs being relatively reasonable, as it is not related to the primary differences between our recommended design and the current design. As such, we see Experiment #1 as the most important research to conduct. However, if two such experiments are conducted, Experiment #2 should probably be conducted prior to Experiment #1 for purposes of knowing how to train the respondents in use of the tablet computers and other equipment.

Experiment #3: Monetary Incentives and Non-Monetary Feedback

Early research on expenditures suggested that small nonmonetary incentives significantly improved the collection of diary information (e.g., Sudman and Ferber 1971). More recent reviews of the existing literature (e.g., Singer 2002) suggest that prepaid monetary incentives increase survey cooperation rates, but the effects of prepaid monetary incentives on consumer diary completion need further investigation. For example, a recent investigation of lotteries on diary completion showed effects of the incentive on cooperation and data quality, but the modes and type of incentive are markedly different from those proposed here (Bonke and Fallesen 2010). Furthermore, how to encourage sampled consumer units to continue to complete the diary during data collection requires greater investigation. This could include electronic follow-ups, telephone calls, or in person visits.

The proposed CE design encourages the use of incentives for initial recruitment and for nonresponse follow-up field work. Here, the proposed experiment is only for the initial recruitment. This experimental factor is proposed to have four levels:[30]

---

[30] We recognize that all of these levels for experimentation may be too low; one could easily modify the levels of incentives, but maintain the design.

Group 1: No incentive: This is the control group, mimicking the current CE no
incentive design.

*Group 2: $10 per consumer unit adult per week promised & $5 per consumer unit
child per week promised:* The U.K. Survey of Food and Living Expenses
offers a 10£ (adult) and 5£ (children) incentive. Because the design
proposed here is longer than other diaries, this experimental factor
compensates each adult in the consumer unit for each week of data
collection. Since this could be a potentially costly design decision, this
factor uses a promised incentive, paid after successful completion of the
week's diary.

Group 3: $10 per week per consumer unit promised: Whether each individual in
the consumer unit needs to be compensated is an empirical question.
This experimental factor uses the same promised $10 per week design as
above, but dramatically reduces costs by compensating at the consumer
unit level, not at the individual level. Comparisons of Group 2 versus
Group 3 allow evaluations of promised CU-level incentives compared to
promised person-level incentive.

Group 4: $10 per week per consumer unit prepaid: This experimental factor also
compensates at the consumer unit level, not at the individual level.
However, it prepays the consumer unit for each week of data collection.
Comparisons of Group 3 and Group 4 permit evaluations of prepaid vs.
promised CU-level incentives.

Practical Insights into the Proposed Design: The proposed design recommends
providing CU's with a monetary token of appreciation for their time and effort
spent completing the diary. How to do this, however, is an open empirical
question given the complexity of the task and the novelty (at least, the 2011
novelty) of the equipment. As such, we have included these four options as what
we think are plausible from a cost standpoint, but also may be meaningful for the
CU.

Estimated Financial and Time Requirements: For simplicity, let us assume that the power calculations from the first experiment hold, and that there will be a 5 percentage point difference in response rates between each of the four groups. [We note that this simplifying assumption may be overly liberal, but it may be a reasonable assumption between the no incentive and the incentivized groups.] If this is the case, then the power calculations indicating n=1565 cases per cell still hold as do the cost calculations for development of the diary and so on.

The primary additional cost arises due to the incentive. There is no extra cost for the 'no incentive group.' The $10 per CU per week prepaid and promised can conservatively be estimated as $10*4*1565 = $62,600 for each treatment. Assuming an average household size of 2.58 persons, we can estimate the total incentives for the per-CU member incentive group of $10*4*1565*2.58=$161,508. Thus, the additional cost for conducting the incentive experiment proposed here is about $287,000. As for timing, we recommend in this experiment for all cases to be released simultaneously. If, as with experiment #1, equipment is limited, a staggered approach outlined there could be done, but will expand the time line accordingly.

Design Features Requiring Further Development

In addition to the methodological studies outlined above, we believe that our recommended design would benefit from the conduct of exploratory work related to several features of our recommended design. For each of these design features, the empirical literature is, in essence, nonexistent, so we have little basis for the recommendation of experimental factors to vary. These exploratory studies would focus on the following design features:

The use of multiple formats of memory trigger diaries and receipts

Further information is needed to understand within-consumer unit technology use. What technology is used by which CU members for what purposes? We know of no national data that currently captures this level of detail (e.g., smartphone ownership with details concerning what

features of the smartphone are used). Are CU members willing to use their personal equipment as part of a data collection effort? Are reimbursements needed for this use? Or are most CU members – even those who are quite technologically savvy – most comfortable with small portable paper diaries for recording expenditures away from home, coupled with maintaining receipts? Furthermore, how comfortable are CUs in providing this information (which may be sensitive or private to some CUs) to BLS electronically? Once BLS has an understanding of the characteristics of CUs and an understanding of the acceptance of requests to use personal equipment and provide copies of receipts, we would recommend small iterative experiments to fine tune the use of particular apps (across various platforms), the development of accessible instructions for the use of those applications, and the development of the most appealing hard copy portable diaries (including the simple diary for those under the age of 15) and receipt retention systems.

The level and frequency of interviewer-respondent interaction over the diary month.

Another factor for which there is little to no experimental literature is the degree to which interviewers should interact with respondents over the period during which they are serving as a diarist. Although most consumer expenditure studies call for an interviewer to follow up with consumer units shortly following the start of the diary period, should the longer diary period prove to be feasible (see Experiment 1) we believe that the BLS should consider flexible protocols for CU and respondent interfaces. These interviewer-CU contacts could include one or more of the following: mid period face to face contacts, telephone contacts, or emails. In part, the degree to which the interviewer will interface with the CUs during the diary period may be determined by the overall design of the field protocol, including the provision of telephone and internet-based help lines for the CU members and the ability to monitor data entry in real

time. We speculate that "best practices" may vary by type of CU, with some units benefiting with frequent reminders and others preferring or requiring little support to complete the task. The lack of literature on this topic leads us to recommend that BLS investigate consumer preferences (e.g., via focus groups) before designing alternative protocols that encompass face to face, telephone, and/or email contacts by the field interviewer.

The incorporation of respondents' financial management and tracking systems. Similar to technology adoption, the use of financial management and tracking software is growing, but we know little about the extent of CU-level penetration nor how or by whom the software is used among members of the CU. To what extent do various financial management systems capture data at the level necessary for use by CE? Should adoption of such software be sufficient to warrant it, can an interface be developed so as minimize redundant data entry for those CUs embracing these various applications? Similar to technology adoption, the use of financial management and tracking software is growing, but we know little about the extent of CU-level penetration nor how or by whom the software is used among members of the CU. To what extent do various financial management systems capture data at the level necessary for use by CE? Should adoption of such software be sufficient to warrant it, can an interface be developed so as minimize redundant data entry for those CUs embracing these various applications? Furthermore, as with receipts, how comfortable are CUs in providing this information in some form to BLS? Here we again think focus groups will be beneficial to understand respondents' willingness to provide access to the information before studying the use of this software at a population level.

The incorporation of feedback mechanisms for responding consumer units.

We have outlined an experimental study to examine the effects of monetary incentives on participation. Beyond incentives, what other feedback or information can BLS provide to responding CUs? What information would be helpful to these CUs? Can BLS move away from a data collection model of little benefit to the CE participants to a model of "data exchange?" We believe that summary reports of expenditures during a month by category may be valued by some households, but 'too much information' or potentially off-putting for others. We also are hesitant to recommend any single feedback mechanism that would imply that BLS is giving financial advice. Here, we believe that extensive qualitative efforts are needed to fully understand what might be most useful to CUs.

V. Considerations concerning Nonresponse

CE response rates have fallen for both the diary and the interview components from roughly 80% to close to 70% over the last decade (Response Rate Status Report, 2011). Although nonresponse rates are not necessarily indicative of nonresponse bias (Groves 2006), a ten percentage point drop in response rates raises concerns about the risk of nonresponse bias of CE estimates. Complicating matters is that the CE has multiple key estimates, including aggregate means, proportions and totals and subclass means, proportions and totals. Nonresponse bias is a property of a statistic, not a survey, and as such, each expenditure estimate will have different nonresponse bias properties.

The changes in the data collection protocol and questionnaires described above will have unknown effects on response rates and on the nonresponse bias properties of the key survey estimates. We hypothesize that with changing the majority of the sample to a single interview and reducing the number of waves for the panel sample, response rates will not drop below their current levels and may even increase. We do not know, however, what the effect of increasing the length of the diary period will be on retention throughout the diary period.

As such, we recommend an active approach to monitoring and adapting to changing conditions during field data collection. In this way, the competing – and sometimes orthogonal – concerns of maintaining high response rates overall and reducing the variability of response rates for important subgroups can be simultaneously examined. This active monitoring does not reduce the need for statistical postsurvey adjustments after the data collection period is over. As such, our recommendations for dealing with unit nonresponse blend monitoring and adjusting data collection protocols during the field period with the statistical weighting adjustments described in the sample design above.

A. Overview of Responsive Design

In a responsive design, data collection efforts are modified conditional on the progress of the field efforts and stability of the survey estimates given the current field protocol (Groves and Heeringa 2006). The changes to the data collection efforts involve a set of prespecified protocol components, such as incentives, shortened questionnaires, experienced interviewers, and so on, specifically designed to bring in sampled consumer unit who differ on the key estimates of interest. Decisions to implement the new protocol are based on close monitoring of field data collection. Thus, to implement a responsive design in the CE, we have to define field effort, key survey estimates, protocol components that may bring in different types of sampled persons, triggers for implementing new protocol components, and a monitoring plan for the field effort itself.

B. Field Effort

The definition of field effort in a responsive design is critical because of the continuous monitoring requirement of a responsive design. Triggers for the implementation of new recruitment protocol features depend on showing that the current protocol no longer yields noticeable gains in response rates or changes in survey estimates over additional exerted effort using the same protocol, a concept that Groves and Heeringa (2006) call 'phase capacity.'

There are many ways to measure field effort. Two of the most common measures are the number of call attempts made to a sampled unit and the number of days that the sampled unit has been in the field. Prior analyses of levels of effort to recruit sampled persons in the CE using the Contact History Instrument (CHI) showed that most sampled units (85%) receive seven or fewer calls in the first wave recruitment (Safir and Tan 2009), receive an average of 4 calls between waves 2 and 3 (Olson 2011), and are in the field for an average of eight days before being resolved (Olson 2011). In general, interviews are resolved faster than noninterviews, but out-of-scope cases are resolved faster

than interviews ('The CEQ Contact History Instrument', 2006). In an analysis of April-December 2005 CE CHI data, 80% of all interviews were completed by the 10th day of the month, and 90% were completed by the middle (15th) of the month.

In the design described above, new sample is released every 30 days. Thus, we recommend implementing new field procedures based on monitoring number of days in the field period, rather than total number of call attempts. We also recommend separately monitoring completion of the household recruitment effort with the in person interview, the diary completion, and panel cooperation.

C. Key Estimates and Data Requirements

Responsive designs preidentify important estimates and field effort indicators to monitor during the course of field data collection. These estimates and indicators can be expansive and overwhelming to monitor. We recommend selecting a small, fixed number of field effort indicators and a small, fixed number of key estimates to monitor overall and across important subgroups.

Overall Field Effort Indicators –Fresh cross-section, Interview component. Field effort indicators are used primarily to gauge costs and monitor field productivity rather than as an error indicator. At a bare minimum, these include overall CU response rates, CU out-of-sample/ineligible rates, CU contact rates, and CU cooperation rates, conditional on contact. We recommend calculating and monitoring both unweighted and base-weighted response, ineligible, contact and cooperation rates daily. Additional indicators could include the proportion of sample units that have been visited at least once, the proportion of housing units that have been visited during 'peak times' (weekday evenings and weekend days), interviewer hours per interview, and so on. With new cross-sections released monthly, variation across months in the day-to-day trends of these field effort indicators will be informative about which protocol features consistently lead to certain field outcomes and which fluctuations are simply noise.

Overall Field Effort Indicators – Fresh cross-section, diary component. We also recommend monitoring overall rates related to the initiation and progress of

the diary. Since the diary will be completed using an internet-enabled tablet, the CU can complete the web survey nightly. To disentangle compliance from lack of purchasing, households will be instructed to submit their survey data each night, even if no member of the household made a purchase that day. The instrument will have an indicator for 'no purchases made today' to allow each household to have something to report.

Roughly 66% of adults are estimated to have broadband internet access in 2010 (Smith 2010; see also Horrigan 2010); the remaining adults will be able to upload their data using wireless data plans.[31] We recommend monitoring the compliance of these two groups separately. For example, relevant field effort indicators to monitor will be the proportion of households who successfully submit data on each day of data collection, the proportion of households who successfully complete a full week of data collection, the number of items reported each day of data collection, change in the number of items reported during data collection over the course of the month, and so on.

Overall Field Effort Indicators – Panel component. The field effort indicators for the panel components are similar to those for the fresh cross-section. An additional source of nonresponse for the panel component is non-location nonresponse (Lepkowski and Couper 2002). Thus, we recommend monitoring the proportion of consumer units that cannot be located in months 7 and 13 to the list of field effort indicators monitored above.

Monitoring of key estimates. Triggers for implementing new protocol components in a responsive design are based on obtaining 'phase capacity' in which the current protocol no longer yields higher response rates and no longer brings in persons who vary on the survey variables of interest. Monitoring of phase capacity of key estimates in the CE is difficult because of the diary component. Thus, we recommend using the global questions about expenditures as proxies of the key estimates to monitor during data collection.

---

[31] A final third group will neither have broadband internet at home and will not be able to have wireless data service.

UWM
UNIVERSITY OF WISCONSIN MILWAUKEE

Abt SRBI

N

Subgroups. Given the lack of information available in real-time for monitoring of key estimates, we recommend monitoring variation in response rates over key subgroups who also vary in levels of expenditures. If all subgroups defined by a particular characteristic have equal response rates, then the risk of nonresponse bias related to that characteristic is eliminated. Previously identified CE correlates of participation include urbanicity, type of PSU, housing tenure (own/rent), housing value, number of rooms in the dwelling unit, region, sex, race, educational attainment, CU size, and CU income (Chopova, et al. 2008). Since the Census 2000 Census 100-Percent Detail File is used as a frame (http://bls.gov/opub/hom/homch16.htm#SampleDesign), we encourage BLS to work with the U.S. Census Bureau to merge auxiliary data from the ACS and/or Decennial Census for purposes of monitoring data collection efforts. Proactive combining of these data, even if the information is out of date for some housing units, will provide more information about characteristics of these subgroups than no information at all.

In absence of micro-level data on each housing unit, BLS should use any and all information available from the ACS and Decennial Census at the Census Tract and Census Block level for monitoring participation rates. Key characteristics that are available from these data sources that seem relevant include median income of the geographic unit, education levels, racial/ethnic composition, renting vs. owning, and single adult households.

Two new sources of data that we recommend be explored are observations by interviewers about the housing unit and the merging on of auxiliary data from administrative databases from companies such as Acxiom, Experian and InfoUSA. CE interviewers traditionally have been asked to make observations about housing tenure (rent/own) for each consumer unit. These observations are combined with respondent reports of owning versus renting and used in post-survey adjustments. We recommend that this information continue to be collected. We also recommend developing a small number of additional observations for interviewers to collect on a small number of other characteristics of the consumer unit that are related to expenditures, such as the presence or

absence of children (Table 5 of the CE Average Annual Expenditures Tables) and estimates of the housing value or income of the CU (Table 2 of the CE Average Annual Expenditures Tables) (Bureau of Labor Statistics, 2010).

Administrative data bases such as credit records created by Acxiom, Experian and InfoUSA provide information on names, addresses, demographics, and consumer purchase behavior. Each list can be obtained for different levels of geography such as zip codes, cities, counties or states, around addresses, or from drawing boundaries on a map. Previous research on administrative databases and survey samples has shown that they cover roughly 75% of addresses in large national samples, that they have large amounts of missing data in the records themselves, and that the quality of the information in these databases on general demographic and socioeconomic characteristics of the household can be quite discrepant from that collected in the surveys themselves (Raghunathan and VanHoewyk no date; see also DiSogra, Dennis and Fahimi 2010). For these reasons, we do not recommend using information from these types of administrative data sets for purposes of reducing burden in the questionnaire. Additionally, in some instances, the survey reports may be more accurate than the information contained in administrative data sets (Sakshaug 2011). However, to the extent that the consumer purchasing behavior in these records is correlated with actual consumer purchasing behavior, then these variables will be more informative than basic demographic information or income information with the CE survey variables of interest.

Montioring subgroups. Whatever auxiliary data are used, they should form the basis for gauging 'representivity' during data collection. One measure of adequate representation of subgroups is variation in response rates over the subgroups (Groves et al. 2008). If there is little variation in response rates across subgroups, then all subgroups are equally represented in the respondent pool. One possible metric is to calculate the coefficient of variation of the response rate over subgroups.

More sophisticated methods include using the auxiliary data to build response propensity models, predicting the probability of obtaining an interview

(or contact or cooperation). Predicted probabilities under these models then summarize the auxiliary information on respondents and nonrespondents. If the variation of these predicted probabilities is small and/or there is reasonable overlap between the predicted probabilities of the respondents and nonrespondents, then the respondents and nonrespondents are similar on the characteristics used to build the propensity models (Rosenbaum and Rubin 1984, 1985).

Alternatively, an R-indicator can be calculated, using a function of these predicted response propensities. R-indicators, or 'representivity indicators' vary from 0 to 1 and indicate how 'representative' the sample is on the characteristics used in the propensity model (Schouten, Cobben and Bethlehem 2009). Alternatively, R-indicators can be built using population information, rather than sample information, to indicate how closely the sample reflects the population distribution on characteristics of interest.

Monitoring Interviewers. Recent work by West and Olson (2010) has shown that nonresponse biases can vary over interviewers and can be of similar magnitude to measurement errors that vary over interviewers. As such, we recommend, at a bare minimum, monitoring response rates across interviewers. Ideally, variation in response rates for a small number of subgroups of interest will be monitored across interviewers. Since the CE does not have an interpenetrated design, we are hesitant in recommending monitoring of key estimates across interviewers as variation across interviewers will also reflect variation across areas and across characteristics of their sample.

D. New Protocol Components

It is currently not known what will be the most effective way to change the recruitment protocol to both increase response rates overall, but also recruit people who vary on levels of expenditures. We suggest three possible options for changing recruitment protocols. To the extent that methodological experiments can occur on field procedures as well as measurement methods, we think that experimental variation on these characteristics will be informative about possible

changes in response rates and nonresponse bias. We do not see these as mutually exclusive, but instead think of them as multiple tools in the arsenal of potential protocol changes.

Shorter Diary Period. The first change to the protocol that may help increase response rates and bring in persons who differ on purchases is a reduction in the length of the diary period. Some nonrespondents may think that a full month of diary taking is unduly burdensome. For these consumer units, then, reducing the length of the diary is a reduction in burden. These respondents will be asked to complete only a two week diary rather than a full month diary. We note that this 'reduction in burden' still increases the amount of information from the consumer unit over the current one week diary. In this way, some 'actual' survey data about relevant consumer unit purchases is obtained, albeit for a shorter time period.

Global Questions Only. Some households may find any recording of expenditures unduly problematic. For these households, when it is clear that they will not participate in the diary, we recommend that the interviewer attempt to complete only the main interview including the global questions, but explicitly mention that the consumer unit will not need to complete the one month diary. Although the global questions will not be used for purposes of estimation, they can be a useful source for nonresponse bias analyses.

Increased Incentives. One of the most consistent recruitment protocol components that increases response rates and potentially brings in a different type of respondent is the use of incentives. Increased incentives for 'end-stage' or 'phrase 2' data collection have been successfully used, for example, in the National Survey of Family Growth (Lepkowski et al, 2010). Here, we recommend offering reluctant households an increase over their original incentive levels. Since our recommendation for incentives is tied to methodological investigation of incentive levels, we recommend a noticeable (e.g., doubling) of the eventually decided upon incentive. For example, if the incentive is $10 per week, then we recommend doubling the incentive for nonresponse reduction purposes to $20 per week.

E. Triggers for New Protocol Components

Field effort and key estimates will be monitored over the course of the one month field period. As a result, a trigger needs to be built that uses the information about the day of the field period after which to implement the new components. Two different approaches can be used – an approach that customizes when a new set of components can be used after a case has been in the field for a prespecified number of days (assuming that all cases are not worked on the same start day) or an approach that universally implements a new protocol for all cases starting on the same day of the field period. For a face to face survey, management of field data collection is likely eased by the latter approach; the former approach can more easily be implemented in a centralized phone facility.

We do now know when during the 30 days 'phase capacity' will be reached. We also expect that 'phase capacity' will be reached at different points for different survey estimates. From analyses of previous CE data (Olson 2010), we suspect that it will occur sometime around day 15 of the field period. However, those analyses tended to look at changes in response rates and not changes in the key estimates. Since there is a 'start-up' time for effectiveness of new protocol components, especially in face-to-face surveys (Olson and Groves forthcoming), we recommend implementing the new protocol components by at least the 20th day of the month. However, the trigger date may be earlier in this protocol.

F. Monitoring System Requirements

Infrastructure is needed to be able to effectively implement a responsive design. In particular, computerized call records with details (day, time, outcome, interviewer, approaches used) about each call attempt made to the consumer unit must be recorded. The CE currently records that information using the Contact History Instrument (CHI). We recommend that this practice continue.

Additionally, an analytic and management system that summarizes and reports on results from call records and disseminates to managers and field staff in real time is also needed. We do not know if the CE currently has this type of infrastructure in place at Census for field data collection. The costs of setting up this kind of system vary tremendously across organizations. Some smaller organizations may simply pay a research assistant to download and analyze these data and disseminate them to relevant field staff; other large organizations custom-build entire reporting systems for this purpose; still other organizations use reporting capabilities present in existing sample management and data collection systems. We do not include the costs of this system – which can clearly be a non-trivial cost – in the estimates for this survey, as the costs for this kind of upfront infrastructure building vary across organizations.

VI. Costs[32]

Our design attempts to balance costs with reductions in errors associated with retrospective recall. We reduce field costs by significantly shifting the field effort from a combination of interview and diary sample to exclusively a diary sample; however we recognize that our design also involves increased levels of "first contacts." In addition, our recommendations include the payment of an incentive, something currently not included in the CE field effort. We leverage these savings against the increased costs associated with increased reliance on technology and incentives.

A.   Field Costs

Cost data for FY 2010 were supplied to us by the National Academy of Sciences; these data suggest that the costs per diary sample case are approximately $250 (total field costs) and range between $270 and $313 per for the quarterly interview survey. Little to no detail on these field costs accompanied the spreadsheet. Due to our lack of understanding with respect to the relationship between the total number of fielded cases and total number of responding sample cases, especially for the diary component, we examined total field costs as a function of sample used for estimation.[33] In FY 2010, total field costs (that is, interviewing labor, mileage, and other direct costs) were reported by the Bureau of the Census to be approximately $21.2 million dollars. One way to view those costs are with respect the data points provided as a result of these expenditures: 7100 diary observations (reported as approximately 14,000 interviews, but we note that these are not independent observations) and 28,000 quarterly interview observations (actually 35,000 quarterly interview observations

---

[32] We question whether we fully understand the costs presented in the spreadsheet. For example, the estimate of $250 per sample unit for the diary interview (which requires at least three in person visits per compliant CU) would suggest CU personal visit costs are under $100 per visit, inclusive of interviewer travel time, mileage, and time spent at the CU.

[33] For example, the data provided indicate that the total CED sample workload in FY 2010 was 14,888 cases. If approximately 7100 CUs provide diary data each year and interviewers contact diarists three times (to enroll them, following week 1 and following week 2) shouldn't the FY 2010 costs be associated with a case workload of 21,000+attempts?

of which 28,000 are used for estimation purposes). Thus, a CU-level observation (regardless of how many face to face interviews or telephone contacts were required to obtain that data) cost approximately $605 in FY2010. Our proposed design calls for an interviewed sample of 33,220 completed CU-level observations per year. Budgeted at the same rate as the effort related to the CEI and the CED leads to an estimated field cost of slightly over $20 million dollars. However, if we budget from the perspective of costs per completed face to face contact and interview (around $325/successful face to face interview) and we are proposing two to three face to face contacts per diarist, we would estimate field costs at between $22.6 million and $32.4 million.

B.   Technology Costs

One of the advantages of incorporating additional technology into the data collection is that technology costs, for a fixed piece of equipment, tend to move in one direction – downward, whereas labor costs tend to move in the opposite direction. In addition, the increased reliance on technology benefits from the societal movement toward greater adoption and penetration across all segments of the population. As a result, costs and adoption rates referenced in this document will quickly be outdated. In light of the dynamics of costs and acquisition rates, we offer this discussion as illustrative of some of the issues that BLS would need to consider in adopting our design.

Our design calls for the placement of two to three pieces of high tech equipment in every CU, for the period in which they serve as a diarist: at most this would consist of a computer tablet, a barcode reader, and a portable scanner. In 2011, the commercial costs for these devices (total) range from about $900 to $1500. These are, of course, retail prices and do not reflect potential savings related to government acquisitions. To support our recommended design, a minimum of 3,000 such sets of units are needed per month; to allow flexibility in the delivery and pick up of the units as well as replacement units and sufficient numbers of units to accommodate geographic dispersion of the sample across the U.S., we would anticipate that the

government would need 6,000 to 8,000 such sets of devices, which at a retail price would cost approximately $6 million to $12 million. Servicing and warranties and/or insurance for the equipment would add further to the costs. Of course, this is cost that can be amortized over multiple years of data collection.

In 2009, slightly more than 60% of households in the U.S. had broadband internet access (Pew Research Center, 2009); in order to provide similar access to all households would require an additional outlay of (once again, based on commercial costs) approximately $25/month/consumer unit.

C.   Incentive Costs

We recommend using an incentive in this design, with a variety of methodological experiments to determine an 'optimal' level of incentives. The least expensive – the promised $10 incentive for each week of completion – would be a total of $1.33 million per year.

D.   Post-Survey Processing Costs

We have no information on current post-survey processing costs in the CEQ or CED. An advantage to a fully web- or computer-based instrument is that the data are more directly captured electronically.

E.   Development Costs

One of the more difficult costs to consider in this redesign effort is the costs associated with development. As part of our research for this contract, we undertook a small feasibility study involving one of the investigators and several graduate students. The feasibility study utilized two pieces of equipment –an iPad 2 and an iPod Touch, free and low cost apps (RedLaser, JotNot Pro, ProOnGo) for barcode reading and scanning of receipts, and one commercially available software package – mint.com. One graduate student –working less than 40 hours –developed a mock protocol of a CE diary (written using Qualtrics) that included several key features: (1) place vs. item specific start to the branching patterns; (2) development of dynamic pick lists for the respondent to

use for repeated purchases; (3) links between diary entries and excel spreadsheet generated from mint.com; (4) links between diary entries and pdfs of receipts created from ProOnGo; and (5) links between the diary entry and the file cabinet web portal of electronic receipts from ProOnGo. The investigator used the devices and protocol for the collection of expenditure data for two weekends in October.

Clearly, the protocol does not represent an instrument one would field for a national study –the point of the feasibility study was to determine if any of the recommendations offered in this report were out of the reach of possibilities for the CE. Our conclusion is that nothing recommended in this report in 2011 is beyond consideration for a national field effort for the future.

Of course, we are not recommending developing a major federal data collection effort with little to no budget. The key costs associated with the redesign include:

- Redesign of the CE Diary, to permit multiple means of entry (direct report, link to excel spreadsheet, link to a scanned receipt, pick from dynamic lists created from user's entries), incorporating "hot keys" and help options that maximize user friendliness and which results in data of high integrity and stability;
- Revision of existing apps and software (or development from scratch) to capture expenditures at the level of detail necessary for CPI use;
- Testing of alternative equipment for durability and reliability;
- Development and testing of communication protocols among devices and between consumer units and web-based portals that maintains the integrity of the data while also protecting the confidentiality of the data;
- Development of training protocols for the respondents and the interviewers.

Redesigns of this magnitude often require multiple years to thoroughly develop and sufficiently test. We forward recommendations from the National Academies' Panel on the Design of the 2010 Census Program for redesign of the

2020 Census (Brown, et al. 2010), including clearly delineating statements of what devices should do (as we have done above), rather than focusing on a single device, as any device currently in use is likely to be sorely out of date in five years. It would not be unreasonable to anticipate development costs of $10 million to $20 million over a two to three year period.

VII. Conclusions and Additional Considerations[34]

The CE data serve many masters, including the CPI (for which detailed information about expenditures is needed and for which the sample size must support sub-national geographic area estimation) as well as outside data users. Optimizing a design for one group runs the risk of limiting the data for different users of the CE data. The design outlined above attempts to meet some of the challenges facing the CE by eliminating reliance on retrospective recall and limiting the source of CPI estimates for expenditures to data obtained from the same CUs utilizing the same design (that is, the same source for food expenditures as well as clothing expenditures as well as mortgage expenditures). The major shortfall of the proposed design is the need to develop synthetic estimates for quarterly and annual expenditures.

There are, of course, an infinite number of ways that one can consider redesigning the CE to meet all of the data users needs; each of these options involves consideration of compromises and rank ordering of priorities. We offer some alternatives below for consideration, in part, to address concerns expressed in the two day workshop held at the end of October 2011.

- **Reduction in burden by splitting requirement for data collection.**
- BLS has, in the past, considered matrix sampling. Of course, doing so eliminates the collection of all expenditure data for each CU. In thinking primarily about burden, an obvious split would be between food expenditures and all other expenditures. However, any matrix approach comes at the cost of redundancies in data collection (e.g., demographic and income information) and requires synthetic estimation to link CUs to have integrated data. The need for increased sample size could be offset by extending the participation from a single month diary to a three month diary period in light of the reduced burden. Furthermore, a greater reliance on digitizing hardcopy receipts or uploading electronic receipts removes

---

[34] This section added post the presentation to NAS, BLS and stakeholders on October 26, 2011.

the need to reduce the items asked for in the questionnaire. **Reduction in burden by limiting the capture of data to "what" or "how much."** CE diary recording is burdensome in part due to the double edged sword of capturing both "what" and "how much." To what extent could CUs capture only the "what" and gather price information from data aggregators? Or for some items, such as food, would it be better to ask CUs to capture the total costs and integrate data from data aggregators to capture the "what?" Once again, reduction in burden could be used to request longer periods of participation, thereby facilitating quarterly estimates within a CU, at the cost of less information.

- **Preserving annual panel data.**

  One of the major concerns with the proposed design is the lack of quarterly and annual data at the CU level (apart from synthetic approaches). To address this, one can consider continuation of a panel that incorporates retrospective recall, a prospective data capture (such as the one proposed) but involving the use of high levels of incentives, or the use of a modified approach to the data collection that attempts to capture the expenditures at a higher level of detail than is needed for the CPI. For example, one could envision a panel for which the request is to maintain financial records similar to those recorded by mint.com users. These data capture expenditures within categories (which could be modified for CE use) but does not capture (in the current commercially available version) detail at the level of interest to CPI. Understanding the demographics of current users of this type of software and why they use the financial management software would be valuable in shaping a data collection effort of this nature. A major limitation of such an approach is that the data would be quite different from the data used to compute the CPI.

- **Using administrative data.**

  Consolidation in retailers provides CE with opportunities to capture data apart from burdening CU members. As we noted, a small number of grocery retailers accounts for over 70% of grocery expenditures in urban areas. But not 100%, and likely quite lower in rural areas. The tradeoff becomes one of mixed data quality – and sources of those data – across CUs as a function of geographic location and socio-economic characteristics.

These suggestions represent but a handful of variations one could consider in a redesign of the Consumer Expenditure Survey. Clearly, any redesign effort must prioritize among data users (with respect to demands) and then work to minimize various sources of error (most notably measurement and nonresponse) within fixed time and resource constraints. We welcome the opportunity to continue a dialogue with the Bureau of Labor Statistics as they wrestle with these issues.

Appendix A
Additional Information on Software Referenced in Paper

Mint.com Personal Finance

This application is based around placing all open checking, savings, and investment accounts into one application. Once this application has been set up and account credentials have been entered, the user of the application can update their account data in real time over a cellular or Wifi network. Transactions are downloaded from associated accounts and can be viewed individually within the application. In addition, the user may input their own transactions at point of sale and categorize them appropriately (cash, check, credit card, etc). Within entering transactions as point of sale, users with GPS-enabled devices (smartphones and some tablets) may input the specific merchant from which they are purchasing items. The application conducts a location search and associates the results with Google Maps to find specific merchant locations in the surrounding area.

In addition to entering and syncing data within a user's account, a user may also create specific budgets for items or expenses that they will incur on a monthly basis. Specific amounts of money can be entered into these budget categories and they are updated accordingly when transaction data is synced with accounts associated with the user. When a specific budget amount is exceeded, there are options to notify the user that they have spent over the amount they had set in a given category. Notifications can range from badges (numerical indication on the application icon when it is not in use) on a smart phone such as iOS or Android devices. Users may also set their preferences to alert them with text messages when they have gone over their set amount. Additionally, these notifications can be set to inform the user when unusual amounts of money or large purchases have been sent or received in one of their accounts.

JotNot Scanner Pro

This application is a multi-page scanner and image enhancement tool for storing documents entered as images from another source. The application functions similarly as commercial scanners (and their software) in that the user has the option of image enhancement and cropping after importing the original document. This application reads text that has been imported from a variety of sources. These include, but are not limited to, receipts, black and white and color documents, and handwriting on blackboard or whiteboards. This application contains a range of image processing options that allow the user to adjust the contrast and tints of backgrounds and image adjustment for black and white and inversion of colors. Also, the user may choose to place a timestamp of the import on to an image after it has been imported for processing.

After a user has imported and processed a document through the application. There are options to adjust the overall file size and resolution of the file. This relates to how the file is to be exported, which there are a range of options to do so. If the user connects to the same Wifi network as a computer, tablet, or mobile device, images can be transmitted over Wifi sharing by connecting to a specified IP address through an internet browser. Images processed through the application may also be exported to applications such as Evernote, Dropbox, Google Docs, and iDisk. Additionally, for a fee of roughly $1 per scan, processed images may be exported by fax directly from the application.

Prizmo

This application is a multi-purpose document scanner that uses optical character recognition (OCR) to recognize typed or written text from the source of the image taken or imported into the application itself. This application does not require an internet connection to function, as all of the OCR support is done through the application and not on a web server. This also means that data is stored confidentially on the device and not online (unless exported to an online storage space). This application has a range of different sources from which an image can come. This includes typed or written text on a document, business cards, bills, and whiteboards. After an image has been taken or imported into the application, the user has several customization options before character recognition scans through the document. These include rotation (left and right), cropping the image down to just labels and prices (if scanning a bill), perspective change, and white balance correction. Users may also select between 10 different languages and the application will cater to these preferences once OCR has begun.

After selecting preferences for the document to be scanned, a document is scanned for individual text characters and the results are filtered into two columns: labels and prices (for bills). Individual items purchased are listed on the left and may be edited to reflect accurate labels for individual items. Prices are also listed and may be edited as well. The user may then also input sales tax and tip percentages associated with the bill being scanned. The resulting output is placed within the category of the scanned item and may then be exported as a full image of the document scanned, or as a .csv file made to import into a spreadsheet. Exporting these files can be done through e-mail, MobileMe, DropBox, Evernote, or CloudApp.

ProOnGo Expense with Receipt Reader

This application allows a user to track transactions and expenses in real time through several different forms. After creating an account, a user may input expense data from a receipt, mileage driven in a car, billable time worked,

income from business transactions, and time expenses from calendar events on the user's device. All expenses are viewable in the expenses tab in the application and a user can individually edit a wide selection of features to determine the proper categorization of the expense. When an expense is entered into the application, the user has the ability to attach an image of the receipt for the expense and add a merchant that is associated with the transaction. In addition to adding this information, the application allows the user to attach the amount of tax paid, split the expense, categorize the transaction based on set categories (in which the user can add their own), and by payment type. The application also supports multiple currencies and exchange rates. Once expenses have been entered, the user has a range of options as to how to display and report the amounts of transactions on graphs and reports. These can be exported to spreadsheet format as well as other expense software (QuickBooks).

Data that is entered into this application is stored on servers hosted by ProOnGo LLC, and, thus, an account is required in order to use the software. Offline storage is an option for this application, but an internet connection is required to access and use reporting data and to sync expenses over multiple devices. In addition to all the features that come with the free version of this application, there is a paid 30 day subscription ($4.99 every 30 days) that a user can sign up for in order to have access to the application web portal (syncing in real time across devices), use of the receipt reader, unlimited exporting of expense reports (XLSX, XML, QuickBooks, among others), and technical support. The receipt reader is capable of automatically extracting merchant data as well as amount paid and date. The application automatically imports this into the user's account and fills out the expense report to be filed. This service can be purchased separately from the premium service in bundles of 30 receipts ($2.99) and 50 receipts ($4.99). In addition to receipt scanning, users may also send in digital receipts (email) for extraction in a similar manner.

References and Reviewed Literature

Abernethy, A.M. (1989). "The Accuracy of Diary Measures of Car Radio Audiences: An Initial Assessment." *Journal of Advertising,* 18:33-39.

Abraham, K., Presser, S., & Helms, S. (2009). "How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States," *American Journal of Sociology*, 114:1129-1165.

Aanensen, D.M., Huntley, D.M., Feil, E.J., al-Own, F., & Spratt, B.G. (2009). "EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection." *PLoS One,* 4: 1-7.

Ahmed, N., Brzozowski, M., & Crossley, T.F. (2009). "Measurement Errors in Recall Food Consumption Data." Unpublished manuscript.

Antenna (2011) "Research Shows 40 Percent of US Consumers with Mobile Phones, Smartphones and Tablets Embrace Mobile Banking." Information retrieved September, 2011 http://www.antennasoftware.com/news/press-releases/6-30-2011-research-shows-40-percent-of-us-consumers-with-mobile-phones-smartphones-and-tablets-embrace-mobile-banking-.

Arens A., Rosch, N., Feidert, F., Harpes, P., Herbst, R., & Mosges, R. (2008). "Mobile Electronic Patient Diaries with Barcode Based Food Identification for the Treatment of Food Allergies." *GMS Medizinische informatik, Biometrie and Epidemiologie*, 4. Open source access at http://www.egms.de/static/de/journals/mibe/2008-4/mibe000073.shtml

Bailey, J., Link, M.W., Bensky, E.N., Vanno, L., Lai, J., Benezra, K., & Makowska, H. (2011). "Can your smartphone do this: A new methodology for advancing digital ethnography." Paper Presented at the 66th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

Baker, M.S., & Oeschger, I. (2009). "Description and Initial Evaluation of a Text Message Based Reporting Method for Marine Recreational Anglers." *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science*,1:143-154.

Banks, J. & Johnson, P. (eds) (1998). *How Reliable is the Family Expenditure Survey? Trends in Incomes and Expenditures over Time.* London: The Institute for Fiscal Studies.

Battistin, E. (2003). "Errors in Survey Reports of Consumption Expenditures." London: The Institute for Fiscal Studies.

Battistin, E. & Padula, M. (2010). "Survey Instruments and the Reports of Consumption Expenditures: Evidence from the Consumer Expenditure Surveys." Centre for Studies in Economics and Finance: Naples, Italy.

Beaudette, D.E. & O'Geen, A.T. (2010). "An iPhone Application for On-Demand Access to Digital Soil Survey Information." *Soil Science Society of America Journal,* 74:1682-1684.

Becker, E., Metsis, V., Arora, R., Vinjumur, J., Xu, Y., & Makedon, F. (2009). "SmartDrawer: RFID-based Smart Medicine Drawer for Assistive Environments." *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments.*

Bernabe-Ortiz, A., Curioso, W.H., Gonzales, M.A., Evangelista, W., Castagnetto, J.M., Carcamo, C.P., Hughes, J.P., Garcia, P.J., Garnett, G.P., & Holmes, K.K. (2008). "Handheld computers for self-administered sensitive data collection: A comparative study in Peru." *BMC Medical Informatics and Decision Making,* 8(11): 1-7.

Blair, J., Menon, G., & Bickart, B. (1991). "Measurement Effects in Self vs. Proxy Response to Survey Questions: An Information-Processing Perspective." In P. Biemer, R. Groves, L.Lyberg, N. Mathiowetz, and S. Sudman (eds) *Measurement Errors in Surveys.* New York: Wiley and Sons.

Bonke, J. and Fallesen, P. (2010) "The impact of incentives and interview methods on response quantity and quality in diary- and booklet-based surveys." *Survey Research Methods* 4: 91-101.

Bound, J., Brown, C., & Mathiowetz, N.(2001). "Measurement Error in Survey Data" in J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, *5*. Amsterdam: North Holland.

Branch, E.R. (1994). "The Consumer Expenditure Survey: a Comparative Analysis." *Monthly Labor Review,* pp. 47-55.

Brattland, J. (2011) "New Research Questions in the Consumer Expenditure Quarterly Interview Survey (April 2011). Memorandum for Jay Ryan, Chief, Division of Consumer Expenditures Surveys, January 4, 2011.

Brown, L, Cohen, M., Cork, D., and Citro, F. (eds) (2010) *Panel on the Design of the 2010 Census Program of Evaluations and Experiments.* Washington: D.C.: National Research Council. Available at http://www.nap.edu/catalog.php?record_id=12865#toc

Bureau of Labor Statistics (2010a). "Global Questions in the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010b). "Interviewing Structure in the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010c). "Introduction for BLS CE Redesign Conference Papers." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010d). "A One Sample Design for the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010e). "Proxy Reporting in the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010f). "Recall Period in the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010g). "Split Questionnaire Methods for the Consumer Expenditure Surveys Program." BLS CE Redesign Conference Paper.

Bureau of Labor Statistics (2010h) "CE Average Annual Expenditures" available at ftp://ftp.bls.gov/pub/special.requests/ce/standard/2010/cucomp.txt and ftp://ftp.bls.gov/pub/special.requests/ce/standard/2010/income.txt

Bureau of Labor Statistics (2009) "Survey Source of Data for Consumer Expnditures Survey Integrated Tables, 209. Available at http://www.bls.gov/cex/ce_2009source.pdf.

Bureau of Labor Statistics (2007). *BLS Handbook of Methods.* Available at www.bls.gov/opub/hom/homtoc.htm.

Bureau of Labor Statistics (2006) "Household Data" *Employment and Earnings.* Available at http://www.bls.gov/cps/eetech_methods.pdf.

Casey, W. (2010). "CPI Requirements of CE."

Chopova, B., Edgar, J., Gonzalez, J., King, S., McGrath, D. & Tan, L. (2008). "Assessing Nonresponse bias in the CE Interview Survey: A summary of four studies."

Cook, S., Sembajwe, R., Geisen, E., and Massoudi, B. (2011) "'Can you see it now? Good': Usability Testing of a Mobile Health Application." Paper presented at the Annual Conference of the American Association for Public Opinion Research. Available at http://www.rti.org/pubs/aapor11_cook_pres.pdf

DiSogra, C., Dennis, M., and Fahimi, M. (2010) "On the Quality of Ancillary Data Availabe for Address-Based Sampling." Paper presented at the Joint Statistical Meetings, Vancouver, CA

Eagle, N., Pentland, A., Pentland, S. & Lazer, D. (2009). "Inferring friendship network structure by using mobile phone data." *PNAS,* 106(36): 15274-15278.

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 US Consumer Expenditure Interview Survey."

Edgar, J. (2011). "2010 CE Methods Workshop Report." BLS Office of Survey Methods Research.

Edgar, J., Davis, J., Spell, S., Verlander, R. & Wetzel, G. (2006). Individual Diary Feasibility Test Final Report. Bureau of Labor Statistics unpublished internal report.

Einav, L., E. Leibtag and A. Nevo (2008). "On the Accuracy of Nielsen Homescan Data." *Economic Research Report Number 69.* Washington, D.C.: U.S. Department of Labor.

Eisenhower, D., Mathiowetz, N.A., & Morganstein, D. (1991). "Recall Error: Sources and Bias Reduction Techniques." In P. Biemer, R. Groves, L.Lyberg, N. Mathiowetz, and S. Sudman (eds) *Measurement Errors in Surveys.* New York: Wiley and Sons.

Experien (2011). National Consumer Studies. Information retrieved September 2011 http://www.experian.com/simmons-research/simmons-consumer-research.html

Faruqui, U. (2010). "Assembling a real-financial micro-dataset for Canadian households." *Journal of Economic and Social Measurement,* 35: 63-84.

Filippucci, C. & Ferrante, M.R. (1997). "Individual Diaries and Expense Documents in the Italian Consumer Expenditure Survey." *Journal of Official Statistics,* 13(2): 113-1221.

Foster, K., Meijer, E., Schuh, S., an Zabek, M. (2011) "The 2009 Survey of Consumer Payment Choice" *Public Policy Discussion Papers* Federal Reserve Bank of Boston.

Fowler, G.A. (2010). "E-Commerce Growth Slows, But Still Out-Paces Retail." *Wall Street Journal.* Available at http://blogs.wsj.com/digits/2010/03/08/e-commerce-growth-slows-but-still-out-paces-retail/.

Fuchs, M. & Busse, B. (2009). "The Coverage Bias of Mobile Web Surveys Across European Countries." *International Journal of Internet Science,* 4(1): 21-33.

Garner, T.I., Janini, G., Passero, W., Paszkiewicz, L., & Vendemia, M. (2006). "The CE and the PCE: a comparison." *Monthly Labor Review,* pp. 20-46.

Goldenberg, K.L., McGrath, D. & 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, May 14-17, 2009.

Gonzalez, J.M., & Eltinge, J.L. (2008). "Adaptive Matrix Sampling for the Consumer Expenditure Interview Survey." *Proceedings of the Survey Research Methods Section of the American Statistical Association.* 2081-2088.

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.M. (2006) "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* Vol 70(5):646-675.

Groves, R.M., & Heeringa, S. (2006). "Responsive Desing for Household Surveys: Tools for Actively Controlling Survey Error and Costs." *Journal of the Royal Statistical Society A* 169:439-457.

Haberer, J.E., Kiwanuka, J., Nansera, D., Wilson, I.B., & Bangsberg, D.R. (2010). "Challenges in Using Mobile Phones for Collection of Antiretroviral Therapy Adherence Data in a Resource-Limited Setting." *AIDS Behav,* 14:1294-1301.

Harris Interactive Inc. Public Relations Research. (2011). *The 2011 Consumer Financial Literacy Survey Final Report.*

Heimovitz, H.K. (2003). *Administrative Records Experiment in 2000 (AREX 2000) Outcomes Evaluation.* Downloaded from: http://www.census.gov/pred/www/rpts/AREX2000_Outcomes.pdf

Henderson, S., Brady, V., Brattland, J., Creech, B., Garner, T. & Tan, L. (2011). "Consumer Expenditure Survey Data Users' Needs."

Henderson, S., Passero, B., Rogers, J., Ryan, J. & Safir, A. (2011). "Consumer Expenditure Survey (CE) Data Requirements."

Horrigan, J. (2010) "Broadband Adoption and Use In America" OBI Working Paper Series No. 1. Washington, D.C.: Federal Communications Commission.

Horsfield, G. (2010). *Family Spending: A report on the 2009 Living Costs and Food Survey.* Office for National Statistics report.

Hosoe, S. (2005). "Capturing Moment of Consumption with Smartphone: Case Study from "Capturing Meal and Snack Consumption Among Japanese Female University Students." *Asia Pacific Advances in Consumer Resarch*, Volume 6: 291-298. Available at: http://www.acrwebsite.org/volumes/display.asp?id=11916

IMS (2011). "IMS Prescription Data." Information retrieved September, 2011 http://www.imshealth.com/portal/site/imshealth/menuitem.a46c6d4df3db4b3d88f611019418c22a/?vgnextoid=a4284c30aaaa0210VgnVCM100000ed152ca2RCRD&vgnextchannel=bc42650204850210VgnVCM100000ed152ca2RCRD&vgnextfmt=default

Jea, D., Balani, R., Hsu, J.L., Cho, D.K., Gerla, M., & Srivastava, M.B. (2008). "Diagnostic Quality Driven Physiological Data Collection for Personal Healthcare." *Engineering in Medicine and Biology Society, 30th Annual International Conference of the IEEE,* pp. 2857-2860.

Jenkins, S., Lynn, P., Jackle, A. and Sala, E. (2004) "Linking household survey and administrative record data: What should the matching variables be?" Available at http://www.iser.essex.ac.uk/files/iser_working_papers/2004-23.pdf

Jones, P., Drury, R., & McBeath, J. (2011). "Using GPS-Enabled Mobile Computing to Augment Qualitative Interviewing: Two Case Studies." *Field Methods,* 23(2): 172-187.

Kew, S.T. (2010). "Text messaging: an innovative method of data collection in medical research." *BMC Research Notes.*

King, S.L., Chopova, B., Edgar, J., Gonzalex, J.M., McGrath, D.E., & Tan, L. (2009). "Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey." *Proceedings of the American Statistical Association Section on Survey Research Methods.*

Kojetin, B.A., & Jerstad, S. (1997). "The Quality of Proxy Reports on the Consumer Expenditure Survey." U.S. Bureau of Labor Statistics.

Kojetin, B.A., & Miller, L.A. (1993). "The Intrahousehold Communications Study: Estimating the Accuracy of Proxy Responses at the Dyadic Level." Paper presented at the annual conference of the American Association for Public Opinion Research, St. Charles, Illinois.

Lai, J., Vanno, L., Link, M., Pearson, J., Makowska, H., Benezra,K., and Green, M. (2011) "Life360: Usability of Mobile Devices for Time User Surveys" *Survey Practice.* Available at http://surveypractice.wordpress.com/2010/02/08/life360/

Lau, S.L., Konig, I., David, K., Barandian, B., Carius-Dussel, C., & Schultz, M. (2010). "Supporting Patient Monitoring Using Activity Recognition with a Smartphone." *2010 7th International Symposium on wireless Communication Systems.* P. 810-814.

Leicester, A. & Oldfield, Z. (2009a). "An analysis of consumer panel data." *Institute for Fiscal Studies Working Paper W09/09*

Leicester, A. & Oldfield, Z. (2009b). "Using Scanner Technology to Collect Expenditure Data." *Fiscal Studies,* 30(3): 309-337.

Lepkowski, J. and Couper, M. (2002). "Nonresponse in the Second Wave of Longitudinal Household Surveys." Pp. 259-273 in *Survey Nonresponse*, R. Groves, D. Dillman, J.Eltinge, and R. Little. (eds) New York: Wiley.

Lepkowski, J., Axinn, W., Kirgis N., Ndiaye, S., Mosher, W. and Groves, R. (2010) "Use of Paradata in a Responseive Design Framework to Manage a Field Data Colleciton." NSFG Paper No. 10-012. Ann Arbor, MI: Population Studies Center, The University of Michigan.

Li, G., Schoeni, R.F., Danziger, S., & Charles, K.K. (2010). "New expenditure data in the PSID: comparisons with the CE." *Monthly Labor Review.*

Li, M., Qian, J.P., Yang, X.T., Sun, C.H., & Ji, Z.T. (2010). "A PDA-based record-keeping and decision-support system for traceability in cucumber production." *Computers and Electronics in Agriculture,* 70: 69-77.

Link, M. (2010). "Nielsen Homescan and Life360 Technologies" Paper presented at the BLS Data Capture Technologies Seminar, March 18, 2010.

Link, M.W., Bailey, J., Bensky, E.N., Lai, J., Benezra, K., & Makowska, H. (2011). "Using Smartphones to Measure World Cup Engagement in South Africa: A Life360 Approach." Paper Presented at the 66th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

Marceau, L.D., Link, C., Jamison, R.N., & Carolan, S., (2007). "Electronic Diaries as a Tool to Improve Pain Management: Is There Any Evidence?" *Pain Medicine,* 8(S3): S101-S109.

Martin, C.K., Han, H., Coulon, S.M., Allen, H.R., Champagne, C.M., & Anton, S.D. (2009). "A Novel Method to Remotely Measure Food Intake of Free-Living Individuals in Real Time: The Remote Food Photography Method." *British Journal of Nutrition*, 101(03): 446-456.

Martinez, S. (2007) "The U.S. Food Marketing System: Recent Developments 1997-2006." *Economic Research Report Number 42.* Washington, D.C:. Economic Research Service.

Massagli, M.P., & Hauser, R.M. (1983). "Response Variability in Self-and Proxy Reports of Paternal and Filial Socioeconomic Characteristics." *The American Journal of Sociology,* 89(2): 420-431.

Mathiowetz, N.A. (2010). Self and Proxy Reporting In the Consumer Expenditure Survey Program. Paper Prepared for the Consumer Expenditure Methods Workshop, Bureau of Labor Statistics, December 7-8, 2010.

Mathiowetz, N.A (1987) "Commentary: Questionnaire Design Advisory Conference" Report retrieved September, 2011 from http://www.bls.gov/cex/geminimaterials.htm

Matulewicz, H.H., Kalb, L., Redel, N., & Cole, N. (2011). "Using Scanner Technology to Learn About Foods Acquired by Low-Income Households." Paper Presented at the 66th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

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.

Miller, L.A., & Tucker, C. (1993). "The Intrahousehold Communications Study: A Typology of Family Cohesion." *Proceedings of the American Statistical Association, Section on Survey Research Methods*

Moore, J.C. (1988). "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." *Journal of Official Statistics,* 4(2): 155-172.

Neter, J. (1970). "Measurement Errors in Reprots of Consumer Expenditures." *Journal of Marketing Research, 7,* 11-25.

Neter, J., & Waksberg, J. (1964). "A Study of Response Errors in Expenditure data from Household Interviews." *Journal of the American Statistical Association*, *59,* 18-55.

Official Statistics of Norway. *Survey of Consumer Expenditure: 1997-1999.*

Olson, K. (2011). "Responsive Survey Design for the Current Expenditure Interview Survey." Final Report prepared under contract DOLBO92J13769.

Olson, K. and Groves, R. (forthcoming). "An Examination of Within-Person Variation in Response Propensity over the Data Collection Field Period." *Journal of Official Statistics.*

Olson, K. Wood, H., & Smyth, J. (2010). Does Providing Respondents with Their Preferred Survey Mode Really Increase Participation? Paper presented at the American Association of Public Opinion Research annual meeting, Chicago, IL.

Oudejans, M. and Christian, L. (2011) "Attention and Usability in Internet Surveys: Effects of Visual Feedback in Grid Questions." *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies.* Marcel Das, Peter Ester and Lars Kaczmirek, (eds). New York: Taylor and Francis Group, pp.215-244

Palen, L., & Salzman, M. (2002). "Voice-Mail Diary Studies for Naturalistic Data Capture under Mobile Conditions." Presentation at Computer Supported Cooperative Work Conference 2002, New Orleans, LA.

Palmblad, M. & Tiplady, B. (2004). "Electronic diaries and questionnaires: Designing user interfaces that are easy for all patients to use." *Quality of Life Research,* 13: 1199-1207.

Patchen, R.H., & Webb, B.M. (2002). "The Future is Now: The Very Latest Findings from the US Market Launch of the Portable People Meter in Philadelphia." *Proceedings of the ESOMAR/ARF Worldwide Audience Measurement Conference.* Cannes: ESOMAR.

Peytchev, A. & Hill, C.A. (2009). "Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations." *Social Science Computer Review,* pp. 1-17.

Pew Research Center (2010). "Home Broadband 2010." Available at http://pewinternet.org/Reports/2010/Home-Broadband-2010.aspx (visited April 20, 2011)

Pew Research Center Internet and American Life Project. (2010). "Who's Online: Demographics of Internet Users" http://www.pewinternet.org/Static-Pages/Trend-Data/Whos-Online.aspx. Accessed May 3, 2011.

Raghunathan, T. and Van Hoewyk, J. (undated) "Disclosure Risk Assessment for Survey Mcirodata" Ann Arbor: The University of Michigan.

Rawlings, A., Ash, S., Swanson, D., Reyes-Morales, S., & Chopova, B. (2010). "Cluster Sampling Research for the Consumer Expenditure Surveys: Integrated Results." Consumer Expenditure Survey 2010 Redesign Research.

Reimers, S., & Stewart, N. (2009). "Using SMS text messaging for teaching and data collection in the behavioral sciences." *Behavior Research Methods, 41,* 675-681.

Rönkä, A., Malinen, K., Kinnunen, U., Tolvanen, A., & Lämsä, T. (2010). "Capturing daily family dynamics via text messages: development of the mobile diary." *Community, Work & Family, 13,* 5-21.

Safir, A. (2011). "Survey Design and Data Collection." Presentation to the CNSTAT Panel on the Consumer Expenditure Survey, February 8, 2011.

Safir, A., & 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.

Sakshaug, Joe (2011) "To Link or Not to Link? Assessing the Quality of Administrative Data for Survey Research" Paper presented at the Washington Statistical Society and the D.C. Chapter of the American Association for Public Opinion Research.

Schaeffer, N.C. (1991). "Conversation with a Purpose—or Conversation? Interaction in the Standardized Interview" Chapter 19 in P. Biemer, R. Groves, L.Lyberg, N. Mathiowetz, and S. Sudman (eds) *Measurement Errors in Surveys.* New York: Wiley and Sons.

Schouten, B, Cobben, F. and Bethlehem, J. (2009) "Indicators for the Representativeness of Survey Response" *Survey Methodology*, Vol 35: 101-113.

Silberstein, A., & 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.

Slesnick, D.T. (2001). *Consumption and Social Welfare. Living Standards and Their Distribution in the United States.* Cambridge: Cambridge University Press.

Smith, A. (2010) "Home Broadband 2010" Pew Research Center's Internet and American Life Project. Accessible at http://www.pewinternet.org/Reports/2010/Home-Broadband-2010.aspx

Solis, H., & Hall, K. (2010). *Consumer Expenditure Survey, 2006-2007.* U.S. Department of Labor and U.S. Bureau of Labor Statistics report.

Statistics Norway (2000) "Survey of Consumer Expenditure 1997-1999" Official Statistics of Norway. Report retrieved from http://www.ssb.no/english/subjects/05/02/nos_c721_en/

Stinson, L., To, N., and Davis, J. (2003) "Creating a "User-Friendly" Expenditure Diary." *Consumer Expenditure Survey Anthology*, 3-7.

Sudman, S. and Bradburn, N. (1973) "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association,* Vol 68: 805-815.

Sudman, S. and Ferber, R. (1971) "Experiments in Obtaining Consumer Expneditures by Diary Methods." *Journal of the American Statistical Association*, Vol.66: 725-735.

Sun, M., Fernstrom, J.D., Jia, W., Hackworth, S.A., Yao, N., Li, Y., Li, C., Fernstrom, M.H., & Sclabassi, R.J. (2010). "A Wearable Electronic System for Objective Dietary Assessment." *Journal of the American Dietetic Association.*

Supermarket News (2010) "Top 75 Retailers in 2009."

Swanson, D. (2011). "CE's Sample Design." Presented at the Consumer Expenditure Survey CNSTAT Panel Briefing, Washington, D.C.

Tourangeau, R. and Smith, T. (1996) "Asking Senseitve Questions: The Impact of Data Collection, Mode, Question Format, and Question Context." *Public Opinion Quarterly* Vol 6: 275-304.

Tremblay, J., Lynch, J., & 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.

Trewin, D. (2006). *Household Expenditure Survey and Survey of Income and Housing, User Guide 2003-04.* Australian Bureau of Statistics. Retrieved September, 2011 http://abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6503.02003-04?OpenDocument.

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.

Tucker, C., & Miller, L. (1993). "The Intrahousehold Communications Study: Family Cohesion and the Level of Knowledge About Expenses." *Journal of the American Statistical Association,* 942.

Villagran, M. (2011). "Methodological diversity to reach patients along the margins, in the shadows, and on the cutting edge." *Patient Education and Counseling,* 82: 292-297.

West, B. and Olson , K. (2010) "How Much of Interviewer Variance is Really Nonresponse Error Variance? *Public Opinion Quarterly* Vol. 5:99-114.

Westat (2011). "Data Capture Technologies and Financial Software for Collecting Consumer Expenditure Data." Final Report Prepared for the Bureau of Labor Statistics.

Whitmore, R., & Kizakevich, P. (2010). "Data Collection Platforms for Integrated Longitudinal Surveys of Human Exposure-Related Behavior." *RTI International.*

Williams, R., & Thomson, E. (1985). "Can Spouses be Trusted? A Look at Husband/Wife Proxy Reports." *Demography,* 22(1): 115-123.

Wind, Y., & Lerner, D. (1979). "On the Measurement of Purchase Data: Surveys Versus Purchase Diaries." *Journal of Marketing Research,* XVI: 39-47.

Yu, P., deCourten, M., Pan, E., Galea, G., & Pryor, J. (2009). "The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries." *International Journal of Medical Informatics,* 78: 532-542.