

MEASURING WHAT WE SPEND

Toward a New Consumer Expenditure Survey

Panel on Redesigning the BLS Consumer Expenditure Surveys

Don A. Dillman and Carol C. House, *Editors*

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. DOLJ109J31331 between the National Academy of Sciences and the U.S. Department of Labor, Bureau of Labor Statistics. Support for the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (award number SES-1024012). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-26575-1

International Standard Book Number-10: 0-309-26575-4

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; <http://www.nap.edu>.

Copyright 2013 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Suggested citation: National Research Council. (2013). *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. Panel on Redesigning the BLS Consumer Expenditure Surveys, Don A. Dillman and Carol C. House, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

**PANEL ON REDESIGNING THE BLS
CONSUMER EXPENDITURE SURVEYS**

DON A. DILLMAN (*Chair*), Department of Sociology and Social and
Economic Sciences Research Center, Washington State University

DAVID M. BETSON, College of Arts and Letters, University of Notre
Dame

MICK P. COUPER, Institute for Social Research, University of Michigan

ROBERT F. GILLINGHAM, Independent Consultant, Potomac Falls, VA

MICHAEL W. LINK, The Nielsen Company

BRUCE D. MEYER, Harris School of Public Policy Studies, University of
Chicago

SARAH M. NUSSER, Department of Statistics, Iowa State University

ANDY PEYTCHEV, RTI International

MARK M. PIERZCHALA, Independent Consultant, Rockville, MD

ROBERT SANTOS, The Urban Institute

MICHAEL F. SCHOBBER, New School for Social Research

MELVIN STEPHENS, JR., Department of Economics, University of
Michigan

CLYDE TUCKER, Independent Consultant, Vienna, VA

CAROL C. HOUSE, *Study Director*

ALICIA JARAMILLO-UNDERWOOD, *Program Associate*

ANTHONY MANN, *Program Associate*

LINDA DEPUGH, *Program Associate*

AGNES GASKIN, *Administrative Assistant*

COMMITTEE ON NATIONAL STATISTICS
2011–2012

- LAWRENCE D. BROWN (*Chair*), Department of Statistics, The Wharton School, University of Pennsylvania
- JOHN M. ABOWD, School of Industrial and Labor Relations, Cornell University
- ALICIA CARRIQUIRY, Department of Statistics, Iowa State University
- WILLIAM DuMOUCHEL, Oracle Health Sciences, Waltham, Massachusetts
- V. JOSEPH HOTZ, Department of Economics, Duke University
- MICHAEL HOUT, Survey Research Center, University of California, Berkeley
- KAREN KAFADAR, Department of Statistics, Indiana University
- SALLIE KELLER, Provost, University of Waterloo, Ontario, Canada
- LISA LYNCH, Heller School for Social Policy and Management, Brandeis University
- SALLY C. MORTON, Department of Biostatistics, University of Pittsburgh
- JOSEPH NEWHOUSE, Division of Health Policy Research and Education, Harvard University
- RUTH D. PETERSON, Department of Sociology (emeritus) and Criminal Justice Research Center, The Ohio State University
- HAL S. STERN, Donald Bren School of Computer and Information Sciences, University of California, Irvine
- JOHN H. THOMPSON, National Opinion Research Center at the University of Chicago
- ROGER TOURANGEAU, Westat, Rockville, MD
- ALAN ZASLAVSKY, Department of Health Care Policy, Harvard University Medical School
- CONSTANCE F. CITRO, *Director*

Acknowledgments

The Panel on Redesigning the BLS Consumer Expenditure Surveys (CE) wishes to acknowledge the support, effort, and expertise of many individuals who contributed to this report and the deliberations behind it. As chair of the panel, I want to begin by thanking my fellow panel members for extraordinary efforts they have made over the past 18 months, and the commitment each gave to this team. The panel members brought a wide range of needed expertise to the table along with a dedication to good science and improving the CE. They consistently provided constructive ideas and assessments under tight timelines. They worked hard together to articulate a balanced framework of conclusions and recommendations that we believe will be helpful to the Bureau of Labor Statistics (BLS). It was a true pleasure working with this group of individuals.

The panel wishes to commend BLS for its commitment to a thoughtful, systematic, and comprehensive multiyear process to redesign the CE. The Gemini Project was initiated by BLS in 2009 and has overseen a wealth of research and information-gathering workshops that formed a base from which the panel launched its own activities. BLS has created a well-organized website that provides easy access to this research and information. Michael Horrigan, associate commissioner, Office of Prices and Living Conditions, is a driving force behind the redesign efforts. He provided the panel with his insight and expectations as well as management support. Working with Michael to steer the Gemini Project and support the panel's activities were Jay Ryan, Consumer Expenditure Division Chief, and Adam Safir, Branch Chief of Research and Program Development Branch. Both Jay and Adam were very helpful throughout the entire process—answering

questions, making staff available for assistance, providing suggestions for workshop speakers and participants, and setting a tone of cooperation with the panel. We want to thank Kathy Downey, who is leader of the Gemini Project. Kathy was the main BLS contact with the panel, so our questions and requests flowed through Kathy. She was well organized and worked hard to assist the panel in its requests. Thanks to Jeanette Davis, who managed the contract. Many other BLS staff stepped forward in many ways to assist the panel by answering questions, participating at panel-sponsored meetings or workshops, and carrying out research that proved very useful to the panel's work. These staff include Rob Cage, Bill Casey, Jennifer Edgar, John Eltinge, Thesia Garner, Steve Henderson, Bill Mockovak, Bill Passero, Dave Swanson, and many others.

The panel held four meetings and two workshops in the process of preparing this report. We would like to thank all participants in these meetings and workshops, and particularly the presenters. The Household Survey Producers Workshop was held in June 2011. At that workshop, overviews of the measurement of consumer expenditures in other countries were presented by Guylaine Dubreuil (Statistics Canada), Giles Horsfield (Office of National Statistics, UK), and Peter Paul Borg (European Union). Presentations describing survey designs that add flexibility in data collection mode were given by Richard Hough (U.S. Census Bureau), Jolene Smyth (University of Nebraska), and Jennifer Wine (RTI International). That session was followed by presenters who discussed survey designs that effectively mix data from multiple surveys and/or external/administrative data to produce estimates—Steve Machlin (Agency for Healthcare Research and Quality), Eileen O'Brien (Energy Information Agency), and Van Parson (National Center for Health Statistics). The fourth session in the workshop presented work by Dave Aune (National Agricultural Statistics Service), Jason Fields (U.S. Census Bureau), and Jane Gentleman (National Center for Health Statistics) on the topic of designs that effectively mix global and detail information to reduce burden and measurement error. Jason Fields (U.S. Census Bureau) and Frank Stafford (University of Michigan) discussed designs that use “event history” methodology. In the final session of the workshop the following presenters provided examples of Diary surveys that effectively utilize technology to facilitate recordkeeping or recall—Justin Bailey (The Nielsen Company), Nancy Cole (Mathematica Policy Research), and Paul Kizakevich (RTI International). Our thanks go to all.

The panel, through the National Academies, competitively selected and commissioned two groups to produce research reports laying out a redesign of the CE. We thank Jennifer Edgar (BLS) for her work with the panel's study director to develop a “statement of work” for this competition. Two contracts were awarded—Westat, and a consortium from the

University of Wisconsin–Milwaukee, the University of Nebraska–Lincoln, and Abt-SRBI. The ideas presented in these reports were important as the panel considered its own recommendations. Both reports presented designs that were innovative and well researched. The time frame given the groups was very short, and both groups met the challenge. The panel used their research and ideas extensively. We would like to give special thanks to both teams: The Westat team—David Cantor (team leader), Brad Edwards, Bob Fay, Abie Reifer, and Sid Schneider; and the consortium team—Courtney Kennedy (Abt-SRBI), Nancy Mathiowetz (team leader, University of Wisconsin–Milwaukee), and Kristen Olson (University of Nebraska–Lincoln). These authors presented their reports at the October 2011 Redesign Options Workshop. The panel was fortunate to also have a number of excellent formal discussants at this workshop. Richard Kulka (independent consultant) discussed the two reports from the view of cognitive and methodological issues, while Chet Bowie (NORC) talked about the issues of implementing change in a large ongoing survey. We were also fortunate to have Mark Lino (Center for Nutrition Policy and Promotion, USDA), Clinton McCully (Bureau of Economic Analysis), and Melvin Stephens (University of Michigan and a panel member) discuss both designs from the perspective of data users. The panel thanks all of these discussants for their insightful comments.

The panel reached out to many users of the CE data. First we wish to extend our thanks to staff at several federal agencies who spoke with panel members about how the CE data are used to help administer programs within their agencies. Thanks to Mary Kate Catlin, Centers for Medicare & Medicaid Services; Cathy Cowan, National Health Statistics Group; Kenneth Hanson, Economic Research Service, USDA; Arnold Katz and Marshall Reinsdorf, Bureau of Economic Analysis; and Geng Li, Federal Reserve Board. The panel held a session at the 2011 CE Data Users Conference, and we would like to thank the analysts who attended this session and passed along their comments on the CE. We would especially like to extend our appreciation to Christopher Carroll (Johns Hopkins University) who attended each of the panel's public meetings and workshops, listening intently and providing insightful comments and suggestions at each venue.

This report was reviewed in draft form by seven individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council (NRC). The independent reviewers provided candid and critical comments that assisted the panel in making its report as sound as possible and ensuring that the report meets the NAS standards for objectivity, evidence, and responsiveness to the study charge. We wish to thank the following individuals for their review of this report: F. Jay Breidt, Depart-

ment of Statistics, Colorado State University; Stephen Cohen, National Center for Science and Engineering Statistics, National Science Foundation; Erik Hurst, The University of Chicago Booth School of Business; Timothy P. Johnson, Survey Research Laboratory, University of Illinois, Chicago; Jean Opsomer, Department of Statistics, Colorado State University; Frank Stafford, Department of Economics, University of Michigan; and Alan R. Tupek, Arbitron, Inc. Their comments were extremely helpful, and the report is much improved because of their efforts. Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Robert Bell (AT&T Labs) and Stephen Fienberg (Carnegie Mellon), who served as Review Coordinator and Review Monitor, respectively. They were responsible for making certain that an independent scientific review of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. The panel was privileged to have these two individuals working on the report, and we thank them for their efforts and insight. The panel also wants to acknowledge the efforts of our editor, Paula Whitacre, who did a great job in making the entire report read smoothly.

The panel would like to show appreciation to the many individuals within the National Academies who provided support and assistance throughout this process. First, we thank Connie Citro, director of the Committee on National Statistics. Connie provided support to the panel throughout the entire process, attending panel members and workshops. She also provided her own insights and recommendations when requested by the panel. Her comments were very helpful.

We also want to express our profound gratitude to the panel's study director, Carol House. Her contribution to the panel's work was enormous, from obtaining needed background information and research for the panel deliberations to drafting the text of final report. The latter was particularly challenging because of the need to bring together diverse input from multiple scientific disciplines and meet the prescribed deadlines. The panel also wants to thank the four program assistants who worked with the panel over its lifespan in setting up meetings and workshops, processing financial paperwork, interacting with workshop participants, preparing agenda books, carrying out web searches, and generally making sure that the entire process proceeded smoothly. So our thanks go to Linda DePugh, Agnes Gaskin, Alicia Jaramillo-Underwood, and Anthony Mann. In addition we thank Kirsten Sampson-Snyder and Eugenia Grohman for their many efforts in assisting the report through the steps required for final publication.

Finally, we recognize the many federal agencies that support the Committee on National Statistics directly and through a grant from the National Science Foundation. Without their support and commitment to improving the national statistical system, the committee work that is the basis of this report would not have been possible.

Don A. Dillman, *Chair*
Panel on Redesigning the BLS Consumer Expenditure Surveys

Contents

SUMMARY	1
1 INTRODUCTION AND OVERVIEW	15
Background of the Consumer Expenditure Surveys, 15	
Context for This Study, 17	
Overview of This Report, 19	
2 THE MANY USES OF THE CONSUMER EXPENDITURE SURVEYS	21
CE Data Provide Critical Input for Calculating the Consumer Price Index, 21	
The CE Provides Data Critical in Administering Government Programs, 25	
CE Data: A Cornerstone for Policy Analysis and Economic Research, 27	
Summary, 36	
3 THE CURRENT CONSUMER EXPENDITURE SURVEYS	37
Design and Implementation, 37	
Cost of the Consumer Expenditure Surveys, 45	
4 THE PANEL'S INVESTIGATION INTO THE ISSUES WITH THE CE	47
Feedback from Data Users, 48	
Panelists' Insight as Survey Respondents, 50	

	Household Survey Producers Workshop: Description and Insights, 51	
	Redesign Options Workshop: Description and Insights, 59	
5	WHY REDESIGN THE CE?	68
	Evidence of Underreporting in the CE, 69	
	Measurement Differences Between the Interview and Diary, 77	
	Sources of Response Error in the Interview Survey, 82	
	Sources of Response Error in the Diary Survey, 92	
	Nonresponse, 101	
	Issues Regarding Nonexpenditure Data, 106	
	Summary of Reasons to Redesign the CE, 107	
6	PATHWAY TO AN IMPROVED SURVEY	109
	Overview, 109	
	Panel's Approach to Design and the Commonalities That Emerged, 113	
	Redesign Prototypes, 118	
	Roadmap to Get There, 163	
	Summary, 186	
	REFERENCES AND BIBLIOGRAPHY	188
	APPENDIXES	
A	Dissent and Panel Response	201
B	BLS Communication of Issues	208
C	Uses of the CE in Administering Federal Programs: Debriefing of Program Staff	215
D	Statement of Work for CNSTAT Competitive Solicitation of Design Ideas	223
E	Household Survey Producers Workshop Agenda	229
F	Redesign Options Workshop Agenda	233
G	Biographical Sketches of Panel Members and Staff	235

Summary

The Consumer Expenditure Surveys (CE) are the only source of information on the complete range of consumers' expenditures and incomes in the United States, as well as the characteristics of those consumers. The CE consists of two separate surveys—a national sample of households interviewed five times, at three-month intervals; and a separate national sample of households that complete two consecutive one-week expenditure diaries. For more than 40 years, these surveys, the responsibility of the Bureau of Labor Statistics (BLS), have been the principal source of knowledge about changing patterns of consumer spending in the U.S. population.

In February 2009, BLS initiated the Gemini Project, the aim of which is to redesign the CE surveys to improve data quality through a verifiable reduction in measurement error with a particular focus on underreporting. The Gemini Project initiated a series of information-gathering meetings, conference sessions, forums, and workshops to identify appropriate strategies for improving CE data quality. As part of this effort, BLS requested the National Academies' Committee on National Statistics (CNSTAT) to convene an expert panel to build upon the Gemini Project by conducting further investigations and proposing redesign options for the CE surveys.

The charge to the Panel on Redesigning the BLS Consumer Expenditure Surveys includes reviewing the output of a Gemini-convened Data User Needs Forum and Survey Methods Workshop and convening its own Household Survey Producers Workshop to obtain further input. In addition, the panel was requested to commission options from contractors for consideration in recommending possible redesigns. The panel was further

asked by BLS to create potential redesigns that would put a greater emphasis on proactive data collection to improve measurement of consumer expenditures. This report summarizes the deliberations and activities of the panel. As summarized below and described more fully in its report, the panel drew four conclusions about the uses of the CE and 16 conclusions about why a redesign is needed. The panel also made 12 recommendations about future directions.

PURPOSES OF THE CE SURVEYS

The CE serves several important purposes. The most visible is for calculating the Consumer Price Index (CPI), one of the most widely used statistics in the United States. Calculating the CPI involves multiple data sources. The CE data provide budget shares (weights) for detailed expenditure categories. Much of this detail is not available elsewhere.

Another important use is to provide data critical for administering certain federal and state government programs. For the continuing administration of many of these programs, the CE is the only continuing source of data with sufficient information on households' demographic characteristics, spending, and income.

In addition, the completeness of the CE in measuring household demographics and consumer expenditures, in combination with repeated measurement over a year for the same households, makes it a cornerstone for policy analysis and economic research. Understanding the differential effects of policies and events on consumer expenditures of all types, and the consequences for people of different ages, races, and ethnicities, sizes of households, and regions, relies upon the CE.

WHY THE CE INCLUDES TWO SURVEYS

The modern version of the CE, with its two independent surveys, was first fielded in 1972–1973. It has been conducted annually since 1980 with the same underlying design concept—different methods of data collection to collect different kinds of data.

The Interview survey was designed to collect expenditures that could be recalled for over three months. The focus was on large expenditures, such as property, automobiles, and major appliances, as well as regular expenditures, such as rent, utility bills, and insurance premiums. The Diary survey, on the other hand, was designed to obtain expenditures for smaller, frequently purchased items.

Over time, however, the Interview survey began to collect information on small, frequently purchased items, while the Diary now collects information on many larger items. Thus, the Interview and Diary now collect

information that allows estimates for certain expenditures to be made from either source, using different question wordings and time periods. BLS uses only data from a single source in its published estimates, selecting the “best” source for each item.

Approximately 7,100 interviews, each of which averages about 60 minutes, are conducted each quarter in the Interview survey, with five interviews for each household. Although most data are collected in household visits, an increasing proportion of the later interviews rely completely or partly upon telephone interviews. One-fifth of the sample is new each quarter, with a corresponding one-fifth of households completing the five-interview sequence. The Diary survey collects usable data from 7,100 households per year, each keeping two one-week diaries. Diary placements occur during 52 weeks of the year, with approximately 273 diaries being completed each week.

Most of the cost is associated with the Interview survey, which produces about 36,000 completed interviews per year, compared to about 14,000 one-week diaries. The “total” data collection cost for the CE surveys in 2010 was \$21.2 million, with the Interview survey costing \$17 million, or about 80 percent of the total.

THE PANEL’S INVESTIGATION

The panel received input from a wide variety of sources. Investigations conducted by the Gemini Project provided critical background. Several panel members themselves use CE microdata. The panel also reviewed published research and held a session at the 2011 CE Microdata Users’ Conference. The panel also studied the complexities of the CPI program and how the CE supports it.

Based on these investigations, the panel makes the following conclusions about the use of the CE. (The numbers represent the location of the conclusion in the full report; thus, more background on the conclusions below is in Chapter 4.)

Conclusion 4-1: The CPI is a critical program for BLS and the nation. This program requires an extensive amount of detail on expenditures, at both the geographic and product level, in order to create its various indices. The CPI is the current driver for the CE program with regard for the level of detail it collects. The CPI uses over 800 different expenditure items to create budget shares. The current CE supplies data for many of these budget shares. However, even with the level of detail that it currently collects, the CE cannot supply all of the budget shares used by the CPI. There are other data sources from which the CPI currently generates budget shares.

Conclusion 4-2: The CPI does not utilize the panel nature of the current CE. Instead the national and regional estimates employed by the CE assume independence of households between quarters on the Interview survey, and independence between weeks on the Diary survey.

Conclusion 4-3: The administration of some federal programs depends on specific details collected from the CE. There are currently no other available sources of consistent data across years for some of these programs.

Conclusion 4-4: Economic researchers and policy analysts generally do not use CE expenditure data at the same level of detail required by the CPI. More aggregate measures of expenditures suffice for much of their work. However, many do make use of two current features of the CE microdata: an overall picture of expenditures, income, and household demographics at the individual household level; and a panel component with data collection at two or more points in time.

Most panel members experienced the CE Interview, the CE Diary, or both as a respondent. These “field” experiences provided broad understanding when connected with insight from top methodological researchers through the Gemini Project’s CE Methods Workshop (December 2010). In addition, the panel studied findings from periodic debriefings of field representatives on how respondents formulate answers (e.g., use of records vs. no records) and the challenges respondents face in answering accurately.

The panel’s Household Survey Producers Workshop (June 2011) was organized around several critical topics, including consumer expenditure surveys in other countries and survey design experiences on other topics and issues. The workshop brought together U.S. and international presenters; university, private-sector, and government-sector experiences; and data collection experiences on a myriad of topics.

The next step was to commission two groups of researchers to develop potential redesigns for the CE surveys. Their proposals encouraged outside-the-box thinking on new collection strategies, technologies, and procedures. The two proposals were presented at a Redesign Options Workshop organized by the panel in October 2011.

Thus, the panel was challenged to bring together the diverse experiences of data users: from those who use it for the CPI to those who study consumer behavior. It was further challenged by the work of statisticians and survey methodologists who design sampling strategies and survey questionnaires to improve data quality in varied situations. In addition, it was challenged by the practical requirements of data collection and new ideas to improve data quality.

WHY THE CE NEEDS TO BE REDESIGNED

The CE surveys are long and arduous. In the Interview survey, the typical respondent answers without being able to consult other members of the household and only infrequently refers to records. The level of detail exceeds what a person can recall for a three-month period. In the Diary survey, respondents are asked to remember to record details of many small purchases and to list each expenditure separately in a complicated booklet.

In addition, consumer spending has changed dramatically over the past 30 years through such things as online shopping, electronic banking, payroll deductions, and greater use of debit and credit cards. Shopping in “big box” stores that sell a huge variety of items challenges people’s ability to recall the amount spent on specific categories of expenditures.

Through comparisons, reported expenditures in both the Interview and Diary surveys tend to be lower than the amounts suggested by the Personal Consumption Expenditures (PCE) data from the Bureau of Economic Analysis. Although there are important conceptual differences between the CE and PCE, the differences suggest both CE surveys underreport consumer expenditures.

Conclusion 5-1: Underreporting of expenditures is a major quality problem with the current CE, both for the Diary survey and the Interview survey. Small and irregular purchases, categories of goods for specific family members, and items that may be considered socially undesirable (alcohol and tobacco) appear to suffer from a greater percentage of underreporting than do larger and more regular purchases. The Interview survey, originally designed for these larger categories, appears to suffer less from underreporting than does the Diary survey in the current design of these surveys.

Estimates derived from the Interview and Diary differ significantly for many expenditure categories in part because many questions are posed in quite different ways. For example, the Interview asks for an estimate of the household’s weekly expense for grocery shopping and then for portions spent for nongrocery items. In contrast, the Diary asks for a listing of individual food items purchased for home consumption during a specific week of the year.

Conclusion 5-2: Differences exist between the *current* Interview and Diary reports of expenditures. Differences in questions, context, and mode are likely to contribute to these differences. The error structures for the two surveys, and for different types of questions in the Interview survey, may be different. Because of these differences we cannot

conclude whether a recall interview or a diary is inherently a better mode for obtaining the most accurate expenditure data across a wide range of items. Both have real drawbacks, and a new design will need to draw from the best (or least problematic) aspects of both methods.

Sources of Underreporting in the Interview

The panel's review suggests underreporting of expenditures may stem from a number of considerations, rather than a single cause. Asking respondents to spend more than five hours over the course of a year answering detailed questions about their expenditures is a substantial burden. The field representative, concerned about the respondent's willingness to agree to additional interviews, may be hesitant to press too hard for accurate recall or the use of records. Under these conditions, it seems likely that the field representative and respondent both benefit from keeping the interview as short and pleasant as possible.

Conclusion 5-3: Motivational factors of both the respondent and field representative appear to negatively influence the quality of the CE Interview data. This leads the panel to the judgment that a changed incentive and support structure for both respondents and field representatives will be needed for a future CE redesign to motivate high-quality reporting and reduce fatigue.

It becomes apparent to Interview respondents that answering "Yes" to a particular question (e.g., "Did you purchase any pants, jeans, or shorts?") leads to being asked a number of detailed, follow-up questions. The respondent is then asked whether they purchased other "pants, jeans, or shorts," and the cycle begins again.

Conclusion 5-4: The current structure of the Interview questionnaire cycles down through global screening questions, and asks multiple additional questions when the respondent answers "Yes" to a screening question. As this cycle repeats itself, a respondent "learns" and may be tempted not to report an expenditure in order to avoid further questions.

Recall of specific detailed expenditures is further complicated because the item may be only one of several items in a single purchase. The diverse ways of purchasing, paying, or authorizing payment and the challenge of connecting specific expenditures to any payment records seem likely to encourage estimation rather than exact reporting.

Conclusion 5-5: The current design of the CE Interview questionnaire makes the cognitive task of recalling expenditures difficult and encourages estimation.

Some questions on the Interview survey are particularly difficult, such as asking respondents to report exact amounts of savings or value of assets now compared to one year earlier and exact dates of purchases for specific items. Even respondents who keep meticulous records may find that their records are not organized to allow honest answers to some questions.

Conclusion 5-6: Some questions on the current CE Interview questionnaire are very difficult to answer accurately, even with records.

It is not possible for many people to recall exact dates and amounts of expenditures over three months. Whereas certain items (e.g., house payments) may not vary and thus can be remembered, others (e.g., clothing and food away from home) may vary dramatically.

Conclusion 5-7: Three months is long for accurate recall of many items on the CE Interview survey. This situation is exacerbated by the ancillary details that are collected about each recalled expense. Errors of omission are likely to occur, and are a contributing factor to the underreporting of expenditures on this survey. Short recall periods, however, may produce more variability in the estimates and provide difficulties for economic research.

Field representatives report the use of records in the interview varies greatly. However, the proportion of respondents who never or only sometimes use records far exceeds the proportion that always or almost always does. Records are used even less when the interview is conducted by telephone.

Conclusion 5-8: The use of records is extremely important to reporting expenditures and income accurately. The use of records on the current CE is far less than optimal and varies across the population. A redesigned CE would need to include features that maximize the use of records where at all feasible and that work to maximize accuracy of recall when records are unavailable.

Field representatives attempt to interview the person most knowledgeable about expenditures. Most interviews do not involve others, and the respondent may not know certain expenditures made by other adult or teenage household members.

Conclusion 5-9: The use of proxy reporting on the CE Interview is problematic, and is a potential cause of underreporting of expenditures.

About one-third of the CE interviews, especially the later ones, are completed by telephone. These interviews result in fewer positive answers to screener questions and do not benefit from an information booklet designed to encourage recall when the field representative visits the household.

Conclusion 5-10: Telephone interviews appear to obtain a lower quality of responses than the face-to-face interviews on the CE, but a substantial part of the CE data are collected over the telephone.

Sources of Underreporting in the Diary

The Diary survey uses a proactive process that involves instructing respondents to report expenditures by all members of the household, asking that they record expenditures daily, and providing detailed instructions on how to complete the diary. However, evidence that important categories of expenditures are less well reported in the Diary than the Interview suggests the full potential of the Diary is not being realized. As with the Interview, several factors may affect the accuracy of Diary reporting.

Diary reporting asks for expenditures in four categories, with each entry asking for multiple pieces of information, placing considerable burden on respondents. Many find it time-consuming and difficult to partition receipts into the requested categories. Motivation to complete the diary appears to decline over the two-week period.

Conclusion 5-11: A major concern with the Diary survey is that respondents appear to suffer diary fatigue and lack motivation to report expenditures throughout the two-week data collection period, and especially to go through the process of recording all items in a large shopping trip.

Field representatives report some respondents see the 44-page diary as too difficult to complete. They are asked by the field representative to collect receipts, to be recorded in the diary during the second household visit.

Conclusion 5-12: A lot of information is conveyed to the diary respondent in a short amount of time. The organization of the diary booklet may result in considerable frustration among some individuals, who feel they cannot master the instructions. They choose instead to collect receipts and leave them for the field representative to enter during the follow-up visit.

The request to record expenditures by day and into broad categories requires respondents to flip pages back and forth as they move between instruction and recording pages. The diary lacks a clear navigational path, and the visual layout makes completing the diary difficult.

Conclusion 5-13: It is likely that the current organization of recording expense items by “day of the week” makes it more difficult for some respondents to review their diary entries and assess whether an expenditure has been missed.

The Diary survey has a short reporting period, which creates concerns regarding the collection of larger and less frequent expense items. Also it is difficult to get a picture of an individual household’s normal spending pattern in only two weeks.

It is not known to what extent respondents seek or are able to obtain expenditures from other household members. Even if others are willing to provide such information, they may not provide it to the respondent in a timely manner.

Conclusion 5-14: Although the diary protocol encourages respondents to obtain information and record expenditures by other household members during the two weeks, it is unclear how much of this happens.

Response Rates Have Declined

Response rates in 2010 for the Interview survey were 73 percent and for the Diary survey, 72 percent. These rates have declined over time, as have response rates to most household surveys. Low response from high-income groups is a concern for both surveys.

Conclusion 5-15: Nonresponse is a continuing issue for the CE as it is for most household surveys. The panel nature of the CE is not sufficiently exploited for evaluating and correcting either for nonresponse bias in patterns of expenditure or for lower compliance in the second wave of the Diary survey. Nonresponse in the highest income group may be a major contributing factor to underestimates of expenditures.

In assessing both the response and nonresponse issues, concerns exist about both the Interview and Diary modes. The panel did not conclude that one mode is intrinsically better or worse. However, it believes that different approaches to the use of both methods have the potential to mitigate these problems.

The ability to link CE data to relevant administrative data sources (such

as IRS data or data on program participation) could provide additional richness for economic research as well as providing potential avenues to investigate the impact of nonresponse on the survey results.

Conclusion 5-16: For economic analyses, data on income, saving, and employment status are important to be collected on the CE along with expenditure data. Aligning these data over time periods, and collecting information on major life events of the household, will help researchers understand changes in income and expenditures of a household over time. Linkage of the CE data to relevant administrative data (such as the IRS and program participation) would provide additional richness, and possibly provide avenues to investigate the effect of nonresponse.

PATHWAY TO A NEW SURVEY

The current detail and requirements imposed by the multiple and divergent CE data uses are difficult to satisfy efficiently within a single design, and the panel believes that tradeoffs must be made. The panel recommends a major redesign of the CE, with the first step to determine priorities among the data requirements of the many uses of the CE so tradeoffs can be made in a planned and transparent manner. Such prioritization is the responsibility of BLS and is beyond what would be appropriate or realistic for the panel to undertake.

Recommendation 6-1: It is critical that BLS prioritize the many uses of the CE data so that it can make appropriate tradeoffs as it considers redesign options. Improved data quality for data users and a reduction in burden for data providers should be very high on its priority list.

Recommendation 6-2: The panel recommends that BLS implement a major redesign of the CE. The cognitive and motivational issues associated with the current Diary and Interview surveys cannot be fixed through a series of minor changes.

The panel's most effective course of action (prior to BLS' priority-setting) is to suggest alternative designs to achieve different prioritized objectives. The panel developed three distinct prototype designs:

- *Design A* focuses on obtaining expenditure data at a detailed level through a "supported journal," a diary-type self-administered data collection with tools that reduce recordkeeping while encouraging the entry of expenditures when memory is fresh and receipts available. Design A also has a self-administered recall survey to collect

larger and recurring expenses. It collects a complete picture of household expenses over six months, with reporting periods varying by expense group.

- *Design B* uses a recall interview coupled with a short supported journal. It provides data for 96 expenditure categories (rather than the more detailed expenses provided by Design A) and collects complete expenditures over an 18-month period in three waves. It builds a dataset particularly useful for economic and policy analysis. This design also involves a small follow-on survey used to help understand measurement errors in the main survey.
- *Design C* incorporates elements of both Designs A and B. It collects the detail of expense items as in Design A while providing a household profile for six months. To do both, it uses a more complex sample design and employs modeling, collecting different information from different households.

The panel wishes to state clearly that evidence on how well each of the proposed prototypes would work is missing. The process of selecting a prototype or components of a prototype should be based not only on BLS' prioritization of goals, but also on empirical evidence that the proposed procedures can meet those goals.

Recommendation 6-3: After a preliminary prioritization of goals of the new CE, the panel recommends that BLS fund two or three major feasibility studies to thoroughly investigate the performance of key aspects of the proposed designs. These studies will help provide the empirical basis for final decision making.

The panel offers the following recommendation that should be viewed in the context of BLS' prioritization of the CE goals.

Recommendation 6-4: A broader set of nonexpenditure items on the CE that are synchronized with expenditures will greatly improve the quality of data for research purposes as well as the range of important issues that can be investigated with the data. The BLS should pay close attention to these issues in the redesign of the survey.

All three designs feature tablet computers with wireless phone cards as an essential ingredient. The report offers guidelines on the development and use of tablets in data collection, but stresses the untested assumptions that must be addressed before proceeding with using this tool. The panel also recognizes some households will need paper instruments. These instruments need to be redesigned to align with the tablets for multimode collection.

Recommendation 6-5: A tablet computer should be utilized as a tool in supported self-administration. However, a paper option should continue to be available for respondents who cannot or will not use a tablet computer. Visual design principles should be applied to redesigning the paper instrument in a way that improves the ease of self-administration and is aligned with the tablet modules.

The panel presents a general roadmap for BLS to follow to complete the redesign of the CE. First, it recommends BLS develop a targeted and tightly focused plan to achieve a redesign within the next five years, a roadmap that should be completed and made public within six months. The Gemini Project is in place to do this.

Recommendation 6-6: BLS should develop a preliminary roadmap for redesign of the CE within six months. This preliminary roadmap would include a prioritization of the uses of the CE, an articulation of the basic CE design alternative that is envisioned with the redesign, and a listing of decision points and highest priority research efforts that would inform those decisions.

Another key element of the prototypes is the use of incentives to motivate respondents to complete data collection and provide accurate data. The panel recommends an appropriate incentive program be a fundamental part of the future CE program. The report provides guidelines for developing an incentive structure, but the details can only be determined with appropriate CE-specific research.

Recommendation 6-7: A critical element of any CE redesign should be the use of incentives. The incentive structure should be developed, and tested, based on careful consideration of the form, value, and frequency of incentives. Serious consideration should be given to the use of differential incentives based on different levels of burden and/or differential response propensities.

The panel had numerous discussions about alternative data sources as a replacement or adjunct to collecting survey data. Although the use of such information at the aggregate or the micro (respondent/household) level holds great promise, the panel also recognized such use is accompanied by risk, particularly from a cost/quality tradeoff perspective. A serious risk and concern is over the continued availability of outside sources over time. The panel decided not to recommend specific external datasets in its three prototypes. However, the panel encourages BLS to continue to explore

administrative data sources for the future and provides general guidelines for doing that.

Recommendation 6-8: BLS should pursue a long-term research agenda that integrates new technology and administrative data sources as part of a continuous process improvement. The introduction of these elements should create reductions in data collection and processing costs, measurement error, and/or the statistical variance and complexity of the CPI estimate. The agenda should address the robustness of new technology and a cost/quality/risk trade-off of using external data.

The panel points to the value of a strong internal BLS research staff. It recommends further development and expansion of their research capabilities in order to respond to the rapidly changing contextual landscape for conducting national surveys.

Recommendation 6-9: BLS should increase the size and capability of its research staff to be able to effectively respond to changes in the contextual landscape for conducting national surveys and maintain (or improve) the quality of survey data and estimates. Of particular importance is to facilitate ongoing development of novel survey and statistical methods, to build the capacity for newer model-assisted and model-based estimation strategies required for today's more complex survey designs and nonsampling error problems, and to build better bridges between researchers, operations staff, and experts in other organizations that face similar problems.

Facing the demands of the immediate redesign of the CE and use of tablet computers, the panel recommends BLS find additional expertise through outside experts and organizations.

Recommendation 6-10: BLS should seek to engage outside experts and organizations with experience in combining the development of tablet computer applications along with appropriate survey methods in developing such applications.

Finally, as described above, all three prototypes propose procedures and techniques that have not been researched, designed, and tested. The prototypes are contingent upon new research undertakings. Much "relevant" background theory and research exist, for which the BLS research program and Gemini Project deserve praise. However, they do not provide enough specific answers for these new options. Considerable investment

must be made in researching elements of the proposed designs, to find specific procedures that are not only workable, but also most effective. These prototypes are not operationally ready—much targeted research needs to be done.

Recommendation 6-11: BLS should engage in a program of targeted research on the topics listed in this report that will inform the specific redesign of the CE.

Recommendation 6-12: BLS should fund a “methods panel” (a sample of at least 500 households) as part of the CE base, which can be used for continued testing of methods and technologies. Thus the CE would never again be in the position of maintaining a static design with evidence of decreasing quality for 40 years.

In summary, the CE performs an extremely important role in helping understand the consumption patterns of American households and more appropriately targeting critical policies and programs. The current CE design has been in place for four decades, and change is needed. The change should begin with BLS prioritizing the many uses of the CE so a new design can most efficiently and effectively target those priorities. The panel offers three prototype designs and considerable guidance in moving toward that ultimate redesign.

1

Introduction and Overview

The Bureau of Labor Statistics (BLS) of the U.S. Department of Labor has tracked expenditures of U.S. consumers for more than a century. This chapter provides background for the Consumer Expenditure Surveys (CE), an overview of BLS' recent efforts to improve the quality of the data collected in that survey, and the context within which this study was framed.

BACKGROUND OF THE CONSUMER EXPENDITURE SURVEYS

The CE is the “only Federal survey[s] to provide information on the complete range of consumers’ expenditures and incomes, as well as the characteristics of those consumers” (Bureau of Labor Statistics, 2011a). BLS has fielded surveys of consumer expenditures for more than 100 years. While initially conducted on a periodic basis, these data have been collected from households continually since the early 1980s. Since their inception, the impetus of these surveys has been to obtain information on changes in the cost of living (Carlson, 1974). In fact, the first two such surveys of the 20th century led to the development of the Cost of Living Index, which was the predecessor to the Consumer Price Index (CPI). Providing budget shares (index weights) for the CPI remains a primary reason for conducting the CE. During the Depression of the 1930s, the use of the survey expanded to include more general economic analysis. During the 1950s, 1960s, and 1970s, data on consumer expenditures were collected approximately once a decade (Bureau of Labor Statistics, 2008).

A modern version of the survey was first fielded in 1972–1973, with

the Census Bureau selecting the sample and conducting fieldwork under contract to BLS. This CE design was the first to highlight its current configuration of two separate surveys (a recall Interview survey and a Diary survey) working in tandem (Bureau of Labor Statistics, 2008).

BLS recognized the need to conduct a survey more frequently, noting that “rapidly changing economic conditions highlighted by the oil crisis in the 1970s illustrated the need for more frequent monitoring of the spending patterns of American consumers. . . . Rapid inflation—in excess of 13 percent from 1979 to 1980—further demonstrated the need for more frequent updates to the CPI budget shares than every decade” (Bureau of Labor Statistics, 2010a, p. 1). This led to changing the CE into an annual survey based on the 1972–1973 design.

The availability of microdata from these surveys opened the door to the investigation of a broad range of important questions, in the public as well as the private domains. As noted by the Bureau of Labor Statistics (1978, p. 1), the 1960–1961 survey “was also valuable in satisfying the growing interest of market researchers, government officials, and private users of data on income, expenditures, and assets and liabilities of American families.” Carlson (1974, p. 1) points out that by the time of the next periodic expenditure survey in 1972–1973, “[non-CPI] uses of the data [had] become increasingly important,” including the evaluation of economic policies, provision of supplemental information for the calculation of the National Accounts data, and market research.

The broad use of the CE for multiple research needs, in addition to the calculation of the budget shares for the CPI, persists to this day. As explained by BLS (Bureau of Labor Statistics, 2011a), “[The CE] is used by economic policy makers examining the impact of policy changes on economic groups, by businesses and academic researchers studying consumers’ spending habits and trends, by other Federal agencies, and, perhaps most importantly, to regularly revise the Consumer Price Index market basket of goods and services and their relative importance.” Jay Ryan (2010), director of the Consumer Expenditure Survey Division in BLS, said in a presentation to the June 2010 CE Data Users’ Needs Forum that the purpose of the CE is “to collect, produce, and disseminate information that presents a statistical picture of consumer spending for the Consumer Price Index, government agencies, and private data users.”

Since the 1980 makeover, BLS has improved the basic survey design of the CE (Bureau of Labor Statistics, 1983, 1997). The most important of these improvements were the conversion of the “Interview questionnaire” to computer-assisted personal interviewing (CAPI) in 2003 and a more “user-friendly” redesign of the Diary form in 2005. Other smaller changes also have been made, often associated with a regular biennial review that can initiate over 100 changes to the questionnaire and survey procedures.

Even so, modifications in the CE design have not kept pace with the changes in how consumers make (and remember) purchases, the reluctance of the public to respond to surveys, and the availability of newer survey methodology and technology.

CONTEXT FOR THIS STUDY

In 2009, perceived decline in the quality of data collected spurred BLS to embark upon a concentrated effort to study and understand the potential sources of error in the CE. Known as the Gemini Project, this multiyear project seeks both to understand the potential causes of the decline in CE data quality and to design improvements that would reduce measurement error. As part of the Gemini Project's efforts, BLS asked the National Research Council, through its Committee on National Statistics, to convene an expert panel. Box 1-1 provides the Statement of Task for the Panel on Redesigning the BLS Consumer Expenditure Surveys (referred to as "the panel" in this report).

Michael Horrigan, BLS associate commissioner for prices and living conditions, addressed the panel at its first meeting on February 3, 2011, providing more specific expectations. He called for flexible recommendations, saying that the "design recommendations should include a menu of comprehensive design options with the highest potential, not one specific all-or-nothing design." He also stated that the "design recommendations should be flexible to allow for variation in program budget, staffing resources and skills, ability of the data collection contractors to implement,

BOX 1-1 Statement of Task

The National Research Council will convene an expert panel to contribute to the planned redesign of the Consumer Expenditure Surveys (CE) by the U.S. Bureau of Labor Statistics (BLS). The panel will review the output of a data user needs forum and a methods workshop, both convened by BLS. It will also conduct a household survey data producer workshop to ascertain the experience of leading survey organizations in dealing with the types of challenges faced by the CE and a workshop on redesign options for the CE based on papers on design options commissioned from one or more organizations. Based on the workshops and its deliberations, the panel will produce a consensus report at the conclusion of a 24-month study with findings and recommendations for BLS to consider in determining the characteristics of the redesigned CE (National Research Council, 2011b).

legal agreements to be obtained (e.g., access to other data sources), etc.” A full statement of this communication to the panel is presented in Appendix B.

Historically, the primary use of the CE data has been for the BLS’ CPI, a Principal Federal Economic Indicator of the United States. The CPI program uses the CE data to produce the budget shares for each of 211 expenditure items, and the CE collects detailed expenditures for more than 800 items used in the construction of the budget shares. In order to produce these budget shares, the CE collects a highly detailed disaggregation of a household’s annual spending. As an additional product from the CE, BLS publishes annual expenditure tables, collapsing the 800+ items into 96 different expenditure categories. BLS also produces microlevel data files for use in basic economic research and policy analysis. The users of the microdata generally are satisfied with these more aggregated categories of spending, but they have other requirements. For example, they want a complete picture of spending, income, and assets for each household in the survey. These users also need data collected on these households at multiple time periods to facilitate investigations of how spending and income change in different conditions.

From a survey design perspective, the uses of the CE have competing requirements. Setting expectations in his original communication with the panel, Michael Horrigan stated that the “CE needs to support CPI needs” and the “CE needs to support other data users as much as possible as long as the design to meet those needs meets the needs of the core CE mission” (see Appendix B). BLS also laid out the *CPI Requirements of CE* in Casey (2010). In May 2011, BLS issued a separate paper entitled *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011), which laid out the comprehensive CE needs beyond those of the CPI. The paper states, “for purposes of this document, the CPI constraints are assumed to be suspended. This is a theoretical exercise, and in no way indicates a lack of support for the CPI program after the survey redesign. This is simply to delineate CPI versus non-CPI requirements for the CE” (Henderson et al., 2011, p. 2).

At a Redesign Options Workshop convened by the panel in late October 2011, the breadth of requirements for the CE stimulated considerable discussion. On November 11, 2011, BLS modified its expectations for the panel’s work:

Therefore, contrary to previous direction to the panel that both the *CPI Requirements of the CE* (William Casey, June 17, 2010) and the *CE Data Requirements* (Henderson, Passero, Rogers, Ryan, Safir, May 24, 2011) collectively form the requirements for the survey, the program managers ask that the panel members treat the *CE Data Requirements* as the mandatory requirements for the survey. The CPI data requirements document

is still helpful in terms of providing larger context for data usage, but are not requirements that the panel's recommendations need to meet. We hope that this relaxation of constraints provides the Panel with greater flexibility in considering their recommended design changes. (See also Appendix B.)

The panel has interpreted this modification to its charge as providing it with greater flexibility in design options to consider. In particular, the panel has considered redesign options that, while supporting the CPI, do not provide the full breadth of detailed expenditures currently supplied by the CE to the CPI. Remaining is the need to provide a complete picture of spending, income, and assets for each household, and to capture data for a constant period of time and at a minimum of twice while the household is in the survey sample. There is also flexibility in these requirements: "The CE regards a complete picture of spending at the CU [consumer unit] level to be a requirement, although by using global questions, imputation, or other methods, it is not required that all expenditures be collected at the same level of detail from each CU" (Henderson et al., 2011).

It is important to note that the panel did *not* interpret this modification in expectations as a statement from BLS management that they have decided that the CE will not continue in the future to support the greater level of detail needed by the CPI. That is still an open question.

OVERVIEW OF THIS REPORT

This report summarizes the work of the Panel on Redesigning the BLS Consumer Expenditure Surveys.

After this brief background in Chapter 1, Chapter 2 describes the many uses of the CE. The chapter notes that the CE has three critical but diverse uses, all of which have great importance for U.S. society: input into the CPI, administration of a diverse array of government programs, and research that provides insight into policy decisions such as the effects of taxes or other economic stimuli.

The current design, implementation, and costs of the CE's two components—the Interview survey and the Diary survey—are explained in Chapter 3.

Chapter 4 summarizes the panel's investigations into issues with the current CE. Through a workshop in June 2011 and other feedback, the panel gathered and carefully considered insights about the CE and explored how other large surveys are conducted. The panel also commissioned the development of two proposals on potential CE redesigns as a starting point to consider new directions. These workshop sessions and the two proposals, which contributed to the panel's conclusions and recommendations, are summarized in this chapter.

Chapter 5 lays out the panel's case about why the CE should be substantially redesigned, noting potential sources of error and respondent burden in both the Diary and Interview surveys. The chapter makes note of underreporting in both versions as compared to other sources of consumer information, most notably Personal Consumption Expenditures (PCE).

Chapter 6 presents three potential redesign options developed by the panel, as well as the panel's recommendations in moving forward. As expressed in these recommendations, the panel urges BLS to prioritize the uses of CE data, to create a roadmap for a redesign, and to conduct targeted research to ensure that any new effort is both workable and effective.

Appendix A contains a dissent statement from three panel members followed by a response by the majority of the panel. Appendixes B through F provide additional background information on various panel activities; and Appendix G provides biographical sketches of panel members and staff.

2

The Many Uses of the Consumer Expenditure Surveys

This chapter emphasizes the importance of this unique federal survey by looking in more depth into the broad spectrum of its use. The chapter begins with a discussion of how the Consumer Expenditure Surveys (CE) are used in the construction of budget shares for the Consumer Price Index (CPI). Subsequent sections highlight many other uses of the CE, including a discussion of the role it plays in the administration of certain federal programs as well as in policy analysis and economic research.

CE DATA PROVIDE CRITICAL INPUT FOR CALCULATING THE CONSUMER PRICE INDEX

In 2002, the National Research Council described the essential role of the CPI:

The Consumer Price Index (CPI) is one of the most widely used statistics in the United States. As a measure of inflation it is a key economic indicator. It serves as a guide for the Federal Reserve Board's monetary policy and is an essential tool in calculating changes in the nation's output and living standards. It is used to determine annual cost-of-living allowances for social security retirees and other recipients of federal payments, to index the federal income tax system for inflation, and as the yardstick for U.S. Treasury inflation-indexed bonds. (National Research Council, 2002)

The Bureau of Labor Statistics (BLS) calculates this index by "observing prices for a sample of goods and services that consumers purchase, and then creating aggregate estimates of price change using average expenditure

budget shares from data that CE provides” (Casey, 2010, p. 1). BLS publishes indexes on a monthly basis for different categories of products and services. Casey (2010) provided an in-depth description of the use of CE data by the CPI program. Most of the particulars included in this section are based on that paper.

The CPI program currently produces four indexes. The CPI-U is the most comprehensive index, measuring price changes for all urban consumers.¹ A second index, the CPI-W, restricts that target population to the subset of urban consumer units in which the majority of income is earned in wage-earning or clerical occupations. A third index, the C-CPI-U, has the same population coverage as the CPI-U. Unlike the CPI-U, however, it uses an index formula that accounts for changes in consumer spending patterns in response to changes in relative prices at all levels of index construction. A fourth index, the CPI-E, is an experimental measure that reflects the spending patterns of urban consumer units in which the reference person is 62 years of age or older.

Types of Data Required by the CPI

Currently, the CE provides the CPI with expenditure data for urban consumer units, along with the demographic information necessary to implement the coverage definitions of the indexes described above.

Demographic Data

For the CPI-U, the CE must (1) allow the identification of urban consumer units and (2) support the construction of subnational CPIs. Additional information is required on sources of income, the percent of income from different sources, and the age of the reference person in the consumer unit, in order to construct the CPI-W and the CPI-E, respectively. Finally, information on the housing tenure of the consumer is necessary for calculating expenditures on the components of shelter cost. Although no other demographic information is required for the current set of CPIs, Casey (2010) states that CPI researchers find additional demographic data useful for constructing other experimental indexes and pursuing other research.

¹The CPI-U does not include the spending patterns of people living in rural nonmetropolitan areas, farm families, people in the Armed Forces, and those in institutions, such as prisons and mental hospitals.

Expenditure Data

For almost all expenditure categories, the CPI requires net out-of-pocket expenses exclusive of any finance charges. The main exception to this rule is the requirement that the CE collect the (implicit) rental value of owner-occupied houses to construct the budget share of the CPI component “Owner’s Equivalent Rent.” Expenditures on major home appliances and certain household maintenance expenses for homeowners are also imputed from the expenditures of renters on these items. This is another reason why housing tenure is a critical demographic variable in the CE. The CPI-U does not require expenditure data for investments, life insurance, interest payment, charitable contributions, or business expenses.

Point-of-Purchase Data

Although the CE currently collects a limited amount of information on where consumers purchase goods and services, the CPI does not currently use any of these outlet data. Rather, the CPI program uses a separate survey, the Telephone Point of Purchase Survey (TPOPS) (Bureau of Labor Statistics, 2011f), to gather this information. However, the CPI program would be interested in expanding the outlet data collection on the CE to provide alternatives to the TPOPS that would be more accurate and better integrated with the expenditure data.

Income Data

The CPI uses information on income (total income; income from wage-earning and clerical worker occupations) to classify each unit in or out of the CPI-W population. Other than this, the CPI does not require information on income or any of its components, including child support and alimony payments.

Geographic Detail

Although the CE survey covers all consumer units in the country, the CPI uses only information on urban consumer units. Regarding geographic breakouts, the top priority of the CPI program is to measure the “All-items, U.S. City Average” index with precision. In order to do that, the current sampling methodology and index construction techniques require the CE to provide reliable, accurate expenditure estimates for 38 geographic areas. However, the publication of indexes for the 38 areas is of secondary concern.

CPIs are published on a monthly basis for the country’s three largest

metropolitan areas (New York, Chicago, and Los Angeles), bimonthly for another 11 metropolitan areas, and semiannually (using six-month averages) for 13 additional metropolitan areas. Separate regional indexes are also published for urban areas in three size classes—metropolitan areas with populations greater than 1.5 million, metropolitan areas with populations less than 1.5 million, and all nonmetropolitan urban areas (separate indexes for nonmetropolitan areas are not available for the northeast and west regions). Because of smaller sample sizes for both consumer expenditures and prices, the expenditure breakdowns for these geographically based indexes are less detailed.

Periodicity

Except for calculation of the C-CPI-U, the CPI program requires only annual expenditure estimates from the CE. Annual expenditure estimates needed to calculate the CPI-U, CPI-W, and CPI-E indexes are estimated by averaging annual expenditure budget shares over two consecutive years. The C-CPI-U, on the other hand, uses the expenditure budget shares from adjacent months to calculate price change between the two months, although information from the prior 12 months is used to reduce variance.

Expenditure Category Detail

The key requirement of the CE from the CPI program—the requirement that is potentially most demanding—is the need for *expenditure detail*. The CPI program uses CE data to calculate expenditure budget shares for 8,018 “elementary indexes.” The 8,018 budget shares are derived by multiplying the 211 item strata by the 38 subnational areas for which budget shares are required (31 areas for the 27 cities for which individual indexes are published, with 3 for the New York Combined Statistical Area [CSA], 2 each for the Chicago and Washington-Baltimore CSAs, plus 7 regions by city-size strata). The current CE sample is too small to support independent estimation of 8,018 budget shares, however, so the budget shares for subnational areas are derived by combining expenditure data specific to the subnational area with expenditure data for a broader geographical area that contains the subnational area using composite estimation. Composite estimation “weights” the subnational-area-specific and broader-area estimates according to their precision. The greater the variance in a subnational-area budget share (the lower the variance of the broader-area budget share), the lower the budget share assigned to the subnational-area budget share in the composite estimate. Moreover, the 27 metropolitan areas for which indexes are published are also part of the more aggregated region-by-size-class indexes.

Consequently, the real constraint for using CE data in the CPI program is the need for national item budget shares of an acceptable precision and enough precision at subnational levels to support an acceptable composite-estimation procedure. It was difficult for the panel to infer exactly what this requirement is.

That CPI requirements are not strictly imposed is reinforced by the fact that the necessary precision to select “entry-level items” is honored in the abeyance: “CE does not currently meet this requirement and CPI must aggregate expenditures to the ELI [entry-level item]-Region level in order to have a large enough sample for each probability” (Casey, 2010, p. 10).

THE CE PROVIDES DATA CRITICAL IN ADMINISTERING GOVERNMENT PROGRAMS

The CE data provide an overall picture of consumer expenditures for the nation. In doing so, the CE provides detailed data on very specific expenses not available elsewhere. Federal agencies and some state agencies use a wide range of these specific estimates to administer important programs. Although far from a complete enumeration, this section describes some important uses of the CE in federal programs. Much of this information was presented at the June 2010 CE Data User Needs Forum (see <http://www.bls.gov/cex/duf2010agendafinl.pdf>) and expanded upon when panel members conducted follow-on discussions. A summary of some of those discussions is in Appendix C.

The CE Provides Important Information on the Cost of Health Care

The Centers for Medicare & Medicaid Services in the U.S. Department of Health and Human Services is responsible for producing the National Health Expenditure Accounts. Among many purposes, these accounts allow for the tracking and projecting of health care spending by businesses, households, and governments. These accounts contribute to the discussion of who ultimately pays for health care in the United States and the burden borne by different sectors of the economy to finance health care into the future. These are critical issues for the country today.

The Health Expenditure Accounts obtain data from a number of different sources, requiring consistent data over time. The CE is the source of private health insurance expenditures paid by households for individually purchased insurance. There is no other available source of consistent data. Additionally, the Health Expenditure Accounts use data from the CE to estimate out-of-pocket expenses for major health services. These include expenses for health services not covered (including deductibles and copayments) by insurance and public programs such as Medicare and Medicaid.

It also includes payments into Health Savings Accounts. The demographics and income data collected in the CE allow analysis of these data by age of head of household and household income. Staff at the Centers for Medicare & Medicaid Services use aggregate estimates published by BLS and also the CE microdata for additional analysis. They use income and asset data from the CE for special analyses (Cowan, 2010).

Taxpayers Have an Easy-to-Use Deduction Based on CE Data

The Internal Revenue Service (IRS) calculates “optional sales tax tables” using CE data. Taxpayers filing a Schedule A have the option to deduct state and local sales tax in lieu of state and local income tax. Many taxpayers, particularly those in states without a state income tax, choose this option. Those taxpayers may keep sales receipts throughout the year and calculate the sales tax they paid. Alternatively, they can use the IRS-supplied “optional sales tax tables” or online sales tax calculator to determine their deduction.

The IRS uses CE data to calculate household estimated sales tax for these tables (and for the online calculator). IRS supplies BLS with state and local taxability data. BLS combines this information with CE data to calculate household-level sales tax estimates. BLS provides these estimates, along with variables such as household income and family size, back to the IRS. The IRS models these data variables to produce the “optional sales tax tables” by household income and family size (Lee, 2010).

CE Data Help Support Child Welfare

The Center for Nutrition Policy and Promotion (CNPP) at the U.S. Department of Agriculture publishes *Expenditures on Children by Families*, an annual report that estimates what it costs to raise a child from birth through age 17, broken down by household income. This publication provides an extremely valuable source of information associated with child welfare. States use it in determining child support guidelines and foster care payments. The CE provides the major source of data for this publication, including child-specific expenditures such as clothing purchased for children. CNPP staff also use CE data on general household expenditures, allocating a proportion of these expenditures to children based on other sources of data (Lino, 2010).

CE Data Contribute to the Measurement of Poverty

In 1995, the National Research Council of the National Academy of Sciences issued a report titled *Measuring Poverty: A New Approach*

(National Research Council, 1995). The report criticized the methodology used to make the official poverty measurement and recommended improved methodology based on CE data. This new Supplemental Poverty Measure uses actual expenditure data for food, shelter, clothing, and utilities to derive poverty thresholds that are compared to measurements of disposable income from the Current Population Survey. This Supplemental Poverty Measure is currently being computed in addition to the historical measure (Short, 2010).

CE DATA: A CORNERSTONE FOR POLICY ANALYSIS AND ECONOMIC RESEARCH

Good policy is created on a foundation of high-quality information about available options and sustained by analyses of whether the policy achieves its intended effect. CE data are used extensively to evaluate policy and conduct applied research on a wide range of issues important to American households. The CE's value is that it provides the "complete picture," tying household demographics to data on the complete range of consumers' expenditures and incomes. It is used extensively by economic policy makers examining the impact of policy changes on economic groups, and by businesses and academic researchers studying consumers' spending habits and trends. This section presents examples of the crucial analysis and research that depend on data from the CE. The intent of this section is to illustrate the breadth and depth of research and policy analysis made possible through CE data.²

Effect of Taxes and Tax Rebates Examined Using CE Data

Effects of Possible Cap and Trade Regulation

The Congressional Budget Office (CBO) is often called upon to project the possible effects of pending regulations. Harris and Perese (2010) did this for the highly publicized and politicized Global Warming Pollution Reduction Program proposed in H.R. 2454. Using the CE data, they predicted how the proposed regulation might affect the purchasing power of households at different income levels. (Their presentation at the BLS Data User Needs Forum was not part of an official CBO projection.) Grainger and Kolstad (2010) pursued a parallel but separate effort to the CBO staff members, also using the CE data. After concluding that these indirect taxes would inequitably affect households at lower income levels, they proposed

²Authors of a number of these studies use the terms *expenditures*, *consumption*, and *spending* somewhat interchangeably.

policy options that could mitigate the regressive distributional effects of a carbon emissions policy.

Impact of Direct Taxes on the Cost of Living

Gillingham and Greenlees (1987) defined a cost-of-living index including direct taxes. They used CE data to approximate the “tax and price index” (TPI) at the household level from 1967 to 1985. On average, the TPI increased much more rapidly than a CPI-type index, but the impact of taxes was highly progressive. They also used the TPI to evaluate alternative methods for indexing the federal tax system and to study an indexed system historically, comparing indexation with the CPI to actual tax policy, a tax system with constant parameters, and an “exact” indexing scheme (Gillingham and Greenlees, 1990). They concluded that (1) the sequence of tax reductions implemented between 1967 and 1985 fell short of mimicking indexation, (2) wealthier households would have benefited relatively more than lower-income households from indexation, and (3) CPI indexation would not have completely eliminated bracket creep.

Effect of Added Gasoline Taxes

West and Williams (2004) looked at potential increases in gasoline taxes and the likely distributional effect of those increases. They used the CE data to incorporate behavioral responses to estimate a demand system that included other goods and services as well as gasoline. They recommended implementing a larger gasoline tax and then using those available funds to reduce labor taxes. Archibald and Gillingham (1981) used CE data to analyze the distributional implications of either gasoline rationing or a tax on gasoline. The use of a model developed by Archibald and Gillingham (1980) implies that the incidence of a tax or the benefit of rationing with a “white market” in coupons would be very progressive.

Effect of Taxes on Charitable Giving

Reece and Zieschang (1985), building on Reece (1979), who also used CE data, used CE data to estimate models of the impact of tax deductibility on the level of charitable giving. They used econometric techniques that addressed the complexity introduced by a progressive step function of marginal tax rates to obtain consistent estimates. They then used the estimated parameters to shed light on the impacts of four alternative tax policies on the level of charitable giving. Their results did not support the proposition that the alternative policies they considered would lead to substantial increases in the level of charitable contributions at the cost of relatively small losses in tax revenue.

Impact of Economic Stimulus Payments to American Households

As part of an economic stimulus program, tax rebate checks were mailed to American households during the summer of 2001. Did households use these rebates in ways that would help stimulate the economy? Exploiting the panel data aspect of the CE, Johnson, Parker, and Souleles (2006) found that households spent roughly two-thirds of their rebate checks during the first six months after receipt. This study was possible because of the addition of questions to the CE to collect information about the amount of the stimulus checks and when they were received.

A similar economic stimulus program was initiated in May 2008. To analyze the 2008 stimulus, questions were again added to the CE about the rebate checks, including a question about what the households explicitly did with the checks. Paulin (2011) found that 49 percent of recipients used the money to pay off debt, while 30 percent reported that they spent the money. Younger recipients were more likely to spend the rebate than were older recipients.

CE Data Lead to a Better Understanding of the American Household

Gender Makes a Difference

Can the relative contributions to running a household by the members of that household be measured? Does gender make a difference in the value of the contribution? De Ruijter, Treas, and Cohen (2005) used data from the CE to “value” some routine domestic tasks, and categorized those tasks as typically “male” or “female.” For example, doing laundry might be a typically “female” task, while mowing the lawn a typically “male” task. They “valued” these tasks by equating their value with the amount households spent when they outsourced those specific domestic services. They also examined how those expenditures differ by living arrangement.

In an examination of the effect of gender on certain purchasing patterns, Kroshus (2008) assessed how much was spent by households on commercially prepared food (as a percent of total food expenditures) by gender and marital status. Not surprisingly, households headed by unmarried men spend a higher percent of their food expenditures on commercially prepared food.

Age Makes a Difference

Fisher et al. (2007) used 20 years of CE data to examine financial characteristics of older adults related to their home. As individuals grow older, their homes become increasingly mortgage-free. Even though this usually means that home equity also increases over this time period, few older

homeowners take out equity loans. CE data provide a unique way to look at generational differences. Paulin (2008) compared young never-married adults in 2004–2005 with similar individuals who responded on the CE two decades earlier (1984–1985). In real dollars, the 21st century young people spent a greater percent of total expenditures on shelter, utilities, and education. They spent less on food, transportation, and apparel than their 1980s counterparts. Health care was relatively unchanged.

Weagley and Huh (2004) used the CE data to look at the dynamics of retirement and near-retirement status on leisure expenditures. Not surprisingly, they found a positive correlation between leisure expenditures with retirement, income, and education.

Race and Ethnicity Make a Difference

The CE data are an ideal source for research on consumption spending as it differs by household racial and ethnic compositions. García-Jiménez and Mishra (2011) examined the demand for meat and meat products and found significant differences among households. Their results showed that white households purchase less meat (especially chicken and seafood products) than do Hispanic households. African American households purchase more pork and chicken than do Hispanics.

Marriage and Cohabitation Make a Difference

Households headed by single mothers, and how their income and consumption changed as a group between 1993 and 2003, were studied by Meyer and Sullivan (2008) using the CE data. The authors reported that income fell sharply (16%) in the first couple of years and then began to rise (17%) over the rest of the decade. For consumption, the authors found a modest (7% to 12%) rise throughout the decade.

Hawk (2011) used the CE data to understand differences in consumption spending between single people and married couples in their twenties. He found that the per capita income of singles was significantly lower than their married counterparts, that married couples were more likely to be homeowners, and that singles spent more per capita on housing, apparel, food, and education. Singles also spent less on health care.

DeLeire and Kalil (2005) examined expenditures on children by the living arrangements of their parents. Using data from the CE, they concluded that cohabiting-parent couples spend less money on education and more on alcohol and tobacco than do married-parent couples. Cohabiting-parent families had spending patterns different from those of divorced single-parent households and never-married single-parent households.

Testing Economic Theories of Consumption Behavior

The Life-Cycle/Permanent Income Hypothesis (LCPIH) is the standard economic framework for understanding household spending and saving decisions over time. While the model provides a number of testable implications, the primary predictions involve how households will choose to consume in response to changes in income. When income changes are anticipated, the model predicts that consumption will *not* change contemporaneously, as households base consumption decisions in each period on expected lifetime wealth as opposed to current income. Unexpected income shocks, which alter lifetime resources, will result in consumption changes. Income changes can also be delineated between transitory and permanent income changes. Transitory income changes (e.g., a single-year windfall or loss) will only have a small impact on consumption as these changes have a small effect on lifetime resources. Permanent income changes, which alter income in all future years, will have a much larger effect on household consumption.

The CE has long been the unique data source that has enabled researchers to test predictions of the LCPIH using a broad set of consumption measures. Attanasio and Weber (1995) found that using microdata containing all expenditure measures for each household dramatically alters the empirical findings of previous tests of the LCPIH. First, the authors found that whereas many prior studies using aggregate expenditure (e.g., national time-series) data yield results inconsistent with the LCPIH, using microdata to create aggregates across households in a way that is consistent with the underlying economic theory results in estimates that are consistent with the model. Second, whereas past studies that had limited consumption measures (in many cases, just food consumption) rejected the LCPIH, testing the model using the full set of household consumption available in the CE could not reject the model.

Subsequent studies using the CE focused on clearly predictable changes in household income to avoid the many potential statistical pitfalls that may arise when predicting income changes for households using econometric methods. Some studies continue to find results that are consistent with the LCPIH. For example, Hsieh (2003) found that Alaskan residents, who receive large, annual oil dividend payments each fall, the amounts of which are pre-announced earlier in the year, do not exhibit a change in consumption upon receiving these payments. However, other studies find estimates that reject the LCPIH. Parker (1999) found that household consumption increases in response to intra-year paycheck increases due to households hitting the maximum annual Social Security tax limit, after which they no longer pay Social Security tax for that calendar year. Souleles (1999) found that household consumption increases in response to income tax refund

receipts, although the refund amounts are known to households before the checks arrive. Stephens (2008) found that household consumption increases once vehicle loans are paid off even though the date and amount of the final payment are known in advance by households. Finally, using the CE Diary data, Stephens (2003) found that households increase their daily nondurable consumption when their Social Security checks arrive, in contrast to the predictions of the LCPIH.

Since the LCPIH models the decisions of individual households, households are able to insure themselves against bad outcomes, such as a job loss or disability, only through their own savings. An alternative model that is an important economic benchmark for understanding the amount of risk that households face is the model of full insurance. In this model, individual households are fully insured against their own household-level changes in income, although aggregate-level income changes will influence household consumption (e.g., a village that pools all of its resources in each year and then redistributes them across all households in the village).

Mace (1991) tested the full-insurance hypothesis by exploiting the panel feature of the CE to regress changes in household consumption on changes in both aggregate consumption and household level “shocks” (e.g., changes in income and employment status). She found mixed evidence in support of this benchmark depending on the choice of empirical specification, although the preponderance of the evidence favors the model. However, Nelson (1994) found that alternative methods of measuring key variables, including using more expansive measures of consumption and employment changes, consistently reject the full insurance model. Attanasio and Davis (1996) focused on the large, observable wage changes that occurred between groups, as defined by education level and year of birth, during the 1980s to test the full insurance model. They concluded that the full insurance model overwhelmingly fails to explain the large between-group changes in consumption found in the CE over the same period.

CE Data Help Measure Well-Being Across Households

Income Inequality Across Households

Heathcote, Perri, and Violante (2010) combined data from the CE, the Panel Study of Income Dynamics, the Current Population Survey, and the Survey of Consumer Finances to conduct a systematic study of cross-sectional inequality in the United States. They found both a continuous and sizable increase in wage inequality over the study period.

Krueger and Perri (2006) investigated welfare consequences of this growing inequality. The CE helped reveal that poor households do not measurably change their consumption in response to lower wages, but in-

stead increase their working hours. The authors then assessed “household welfare” consequences using a number of techniques. They concluded that about 60 percent of U.S. households face welfare losses, with the size of those losses ranging from 1 to 6 percent of lifetime consumption for different groups.

Life-cycle models of variability in household savings and wealth accumulation (with comparable socioeconomic configurations) have ascribed cause to such factors as risk aversion, preferences for work or leisure in later life, and income replacement rates. Bernheim, Skinner, and Weinberg (2004) used data from the CE and the Panel Study of Income Dynamics to evaluate these conclusions. Instead, they found the “empirical evidence therefore casts doubt on theories that rely on differences in relative tastes for leisure, home production, or work-related expenses to explain the variation in wealth at retirement” (Bernheim, Skinner, and Weinberg, 2001, p. 854). The authors concluded that these factors appeared to be outside the context of the life-cycle model.

Understanding Poverty and How to Measure It

Fisher et al. (2009) examined the financial well-being of households of older Americans. They first distinguished between the notions of “income” poor and “consumption” poor. The authors emphasized that it is important to understand these two poverty definitions and the populations they imply in order to effectively measure the success of various poverty programs. Using 20 years of CE data, they reported that the measure of “poverty” is cut by one-fourth if that measurement uses both income and consumption. Older households that are white, homeowners, and married and have a high school diploma are more likely to be “poor” using only the income definition and not the combined definition. That is because they have sufficient assets to raise consumption above the poverty threshold.

Potential Nutritional Barriers in Poor Families

Do some households have to choose between paying heating bills and buying food? Bhattacharya et al. (2003) used the CE to track expenditures on both food and home fuels. They found that households (both rich and poor) had to increase expenditures on home heating during particularly cold periods. The difference: Poor families decreased their expenditures on food by about the same amount as they increased expenditure on home fuels, but richer families made no change in food expenditures during these same periods. The authors concluded that social programs need to understand this phenomenon and provide special assistance during cold-weather periods.

Stewart, Blisard, and Jolliffe (2003) concluded that low-income households spend less on fruits and vegetables than other households. They based this conclusion on a study of expenditures on fruits and vegetables and how these expenditures correlate with income. More surprisingly, these same households do not purchase more fruits and vegetables when they have a positive change in income.

Health Care Expenditures

What do households purchase when they do not purchase health insurance? Levy and DeLeire (2008) asked this question and used CE data to try to answer it. They found that households without health insurance spend more (compared to insured households) on things such as housing, food, alcohol, and tobacco. The authors raised the possibility that these households may be uninsured because they spend a greater percentage of total income on basic needs.

Does Medicare eligibility reduce out-of-pocket health care expenditures for those individuals? Have those expenses been changing over time? Duetsch (2008) looked at out-of-pocket health care expenditures of persons whose age was 55–64 (Medicare eligibility is 65) and those 65–74. Using CE data from 1985, 1995, and 2005, she found health care expenses (as a percent of total expenses) increased over those 20 years, but not consistently in real dollars. Between 1985 and 1995, the younger (ineligible) group's health care expenditures decreased 27 percent, while the older (eligible) group's expenditures decreased by 18 percent. Between 1995 and 2005, the younger group's health care expenses rose by 22 percent and the expenses of the older group rose by 9 percent. In both decades, the older group spent more overall on health care than the younger group.

CE Data Are Used to Examine Credit and Debt in American Households

CE data are a tool for examining credit constraints and their effect. Ekici and Dunn (2010) examined credit card debt in its relationship to consumption. They used a monthly survey of credit card use to impute credit card debt into the CE data. They found a negative correlation between debt and change in consumption. Specifically the authors showed that a \$1,000 increase in credit card debt leads to a 2 percent decrease in consumption growth. They also examined credit card debt by various household characteristics.

Grant (2007) investigated whether lower borrowing rates of some groups were related to credit constraints or lower demand. He estimated credit constraints and showed how these constraints differ by various household characteristics. His work found that households headed by

young, college-educated individuals were the most credit constrained. He also found that an observed lower level of borrowing among African American households appears to be affected by demand rather than credit constraints.

CE Provides a Tool to Examine Industry-Specific Markets and Special Topics

The comprehensive nature of the CE data allows for analysis that is targeted to expenditures within specific markets.

Transportation Expenses

The probability of leasing a car is increased for households that are older, white or Hispanic, college educated, living in the Northeast and Midwest, living in a large Metropolitan Statistical Area, not having teenagers, and having a higher income. These results were found by Fan and Burton (2005) as they looked at the demographics that lead to a decision to “buy or lease” an automobile. However, the authors indicated that these effects are diminished when one controls for the vehicle characteristics.

Are communication expenses in some way a substitute for transportation expenses? Models developed by Choo, Lee, and Mokhtarian (2007) showed that these two areas of expenditure have both substitution and complementary effects.

Expenditures on Technology

Yin, DeVaney, and Stahura (2005) built a conceptual model using CE data to estimate the amount of money households are likely to spend on computer hardware and software. They then provided implications for consumers and policy makers. Hong (2007) used the CE data to examine the “substitution” relationship between expenditures on the Internet and other entertainment goods. He found Internet expenses have an effect on expenditures of recorded music.

Charitable and Political Giving by Households

There is a U-shape relationship between charitable giving and household income, with households at both the lower and higher income ranges giving a higher percentage of their income to charity than middle-income households. James and Sharpe (2007) found this result as they used CE data to examine the distribution of charitable giving by household income. The authors found that the charitable givers in the lower income ranges

are proportionately older, low-income but higher asset households. Dehejia, DeLeire, and Luttmer (2007) found that individuals who contribute to religious organizations are better able to insure consumption against income shocks. James (2009) examined the characteristics of households that make political contributions. He used a decade of CE data from 1995 to 2005. His analysis showed that political contributions were positively associated with income, wealth, education, and well-being. Political giving was negatively correlated with being a single female and being nonwhite.

SUMMARY

For over a century, the collection of consumer expenditures on the CE and its predecessor surveys has played an irreplaceable role in understanding the market basket of goods and services that consumers purchase. While providing budget shares for the CPI remains a vital reason for the collection of consumer expenditures, a number of prominent uses of these data have emerged since the inception of these surveys. When contemplating revisions to the CE, it is important to remember that the CE has three critical but diverse uses, all of which have great importance for U.S. society: the CPI, the administration of a diverse array of government programs, and research that provides insight into policy decisions such as the effects of tax or other economic stimuli.

3

The Current Consumer Expenditure Surveys

This chapter describes the two components of the Consumer Expenditure Surveys (CE)—the Interview survey and the Diary survey. The Panel on Redesigning the BLS Consumer Expenditure Surveys identified some limitations associated with these two surveys, and these issues are presented in Chapter 5.

DESIGN AND IMPLEMENTATION

As noted in Chapter 1, the current CE are based on the design of 1972–1973 predecessor surveys (Bureau of Labor Statistics, 2008). The CE consist of two different surveys of U.S. households conducted independently, the Interview survey and the Diary survey. Each of these surveys is designed to represent the total U.S. civilian noninstitutional population, using the 2000 Census 100-Percent Detail File augmented by new construction permits and through coverage improvement techniques.

In both surveys, the sampled unit consists of:

- (1) all members of a particular housing unit who are related by blood, marriage, adoption, or some other legal arrangement, such as foster children; (2) a person living alone or sharing a household with others, or living as a roomer in a private home, lodging house, or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more unrelated persons living together who pool their income to make joint expenditure decisions. Students living in university-sponsored housing are also included in the sample as separate consumer units. (Bureau of Labor Statistics, 2008, p. 2)

This report refers to this sampled unit or consumer unit as a *household*.

The two surveys are sampled and conducted independently. Expenditure estimates are made independently from the two surveys for various purchased items (goods and services). The concept of having two distinct surveys is fairly simple. The Interview survey was “designed to collect data on the types of expenditures respondents can be expected to recall for a period of 3 months or longer. In general, expenditures reported in the Interview Survey are either relatively large, such as for property, automobiles, or major appliances, or occur on a fairly regular basis, such as for rent, utility bills, or insurance premiums” (Bureau of Labor Statistics, 2008, p. 2). On the other hand, the Diary survey was designed “to obtain expenditure data on small, frequently purchased items, which are normally difficult to recall. These items include food and beverage expenditures, at home and in eating places; housekeeping supplies and services; nonprescription drugs; and personal care products and services” (Bureau of Labor Statistics, 2008, p. 3).

In reality, there is considerable overlap in expense items collected on the surveys, increasing the total response burden. The Interview survey, reaching beyond its original design, also collects information on small, frequently purchased items that may be difficult to recall. For example, it collects expenses for prescription medication, fresh flowers, sewing notions, and the full range of clothing and shoes. It also collects average monthly cost of gasoline and the average weekly cost of buying food in a grocery store. Similarly, the Diary survey, with its open listing sheets, collects many larger items identified as appropriate for the Interview survey. For example, it collects information on the purchase of dishwashers, china, jewelry, and vehicles (Bureau of Labor Statistics, 2011b,c).

The final Bureau of Labor Statistics (BLS) estimate presented in the Consumer Expenditure Survey Integrated Tables for any particular item is based on an estimate from one or the other of these surveys. Creech and Steinberg (2011) describe how the survey source is selected for each published item, and Bureau of Labor Statistics (2009a) provides examples of which source was used for different items for the Consumer Expenditure Survey Integrated Tables for 2009. Some choices are fairly obvious. New refrigerators and living room chairs were estimated from the Interview survey. The Diary survey was used to estimate expenditures for “wine consumed at home” and “lunch at a fast food restaurant.” However, the choice is not always intuitive. For example, bedroom linens were estimated from the Diary survey, while curtains and draperies were estimated from the Interview survey. Under the grouping of “housewares,” silver service pieces were estimated from the Diary survey and other service pieces from the Interview survey. Luggage was estimated from the Diary survey, while smoke alarms were estimated from the Interview survey.

BOX 3-1
Web Links to CE Survey Documents

Interview CAPI instrument (2011): <http://www.bls.gov/cex/capi/2011/cecapihome.htm>

Interview Survey Information Booklet 2011: http://www.bls.gov/cex/current/i_info_book.pdf

Diary Survey Form 2005-10: <http://www.bls.gov/cex/csx801p.pdf>

Computer Assisted Diary Household Characteristics Questionnaire 2011-12: <http://www.bls.gov/cex/ced/2011/cedhome.htm>

Diary Survey Information Booklet 2011-12: http://www.bls.gov/cex/current/d_info_book.pdf

The survey questionnaires ask for dollar amounts for services and goods purchased by a household member during the prescribed reference period. They exclude all business-related or reimbursed expenditures.

Survey documents for the CE can be found on the BLS website. Box 3-1 provides the specific links.

Design and Implementation of the Interview Survey

Sampling Frame and Sample Size for the Interview Survey

The sample for the Interview survey begins with a selection of 91 area-based primary sampling units (PSUs). The PSUs may be individual counties, groups of counties, or “core-based statistical areas” (CBSAs) identified by the Census Bureau. Of these 91 PSUs, 21 are metropolitan CBSAs with over 2.7 million people, while 16 are considered “rural.” The rest fall somewhere in between (Bureau of Labor Statistics, 2012) as indicated in Box 3-2.

Each year the Census Bureau selects approximately 15,000 addresses from these PSUs for contact on the Interview survey using the augmented 2000 Census 100-Percent Detail File. The bureau uses a rotating panel design with sampled households contacted quarterly for five quarters. This process results in a usable sample size of approximately 7,100 interviews per quarter.

BOX 3-2
Classification of Primary Sampling Units (PSUs)

The 91 PSUs used in the CE sample are classified into four categories:

1. 21 "A" PSUs, which are metropolitan core-based statistical areas (CBSAs) with a population over 2 million people
2. 38 "X" PSUs, which are metropolitan CBSAs with a population under 2 million people
3. 16 "Y" PSUs, which are "micropolitan" CBSAs, defined as areas that have at least one urban cluster of at least 10,000 but less than 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties
4. 16 "Z" PSUs, which are non-CBSA areas, and are often referred to as "rural" PSUs

SOURCE: Bureau of Labor Statistics (2012, p. 5).

Implementation of the Interview Survey

Figure 3-1 displays the flow process of the Interview survey. Work on the Interview survey begins each month with one-third of the quarterly sample. The Interview survey had an overall response rate of 73 percent in 2010. (See "Comparison of Response Rates" in Chapter 5, for a discussion of how these rates are calculated.) Households receive a pre-survey notification letter. The Interview survey is designed for collection through an "in-person" visit by a field representative, and most data are collected in this fashion. However, field representatives are allowed to fall back to a telephone interview and often do. Of completed cases in 2010,¹ approximately 17 percent were completed entirely via the telephone, and an additional 48 percent were completed in part over the telephone. The relative number of cases completed over the telephone increases over later phases of the survey. Of interviews completed in quarter 1 (initial interview), only 2 percent were interviewed entirely via telephone and an additional 34 percent had some data collection over the phone. By quarter 5 in the rotation, 22 percent of completed interviews were conducted entirely over the telephone, and an additional 50 percent had some data collected over the phone. Whether via personal visit or over the telephone, the field representative uses a computer-assisted personal interviewing (CAPI) instrument.

¹Data on interviews by mode for 2010 come from an internal spreadsheet of costs provided to the panel by BLS.

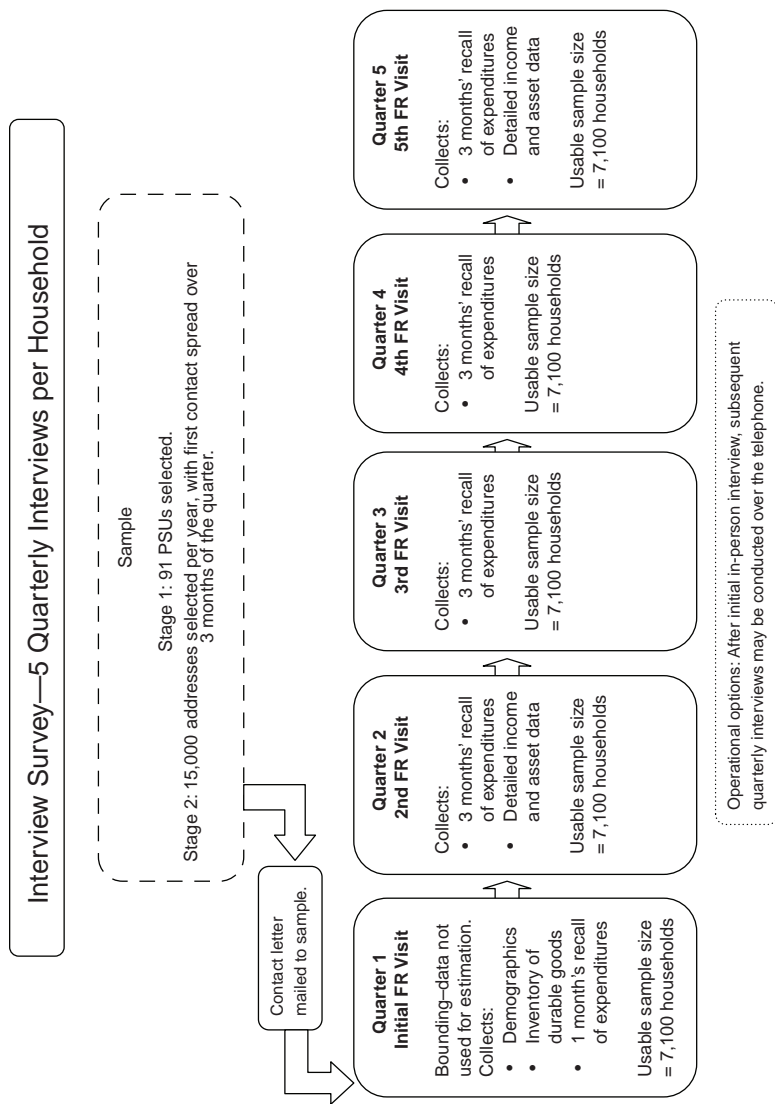


FIGURE 3-1 Process flow of the Interview survey.
 NOTE: FR = field representative, PSUs = primary sampling units.
 SOURCE: Panel designed flowchart based on documentation (Bureau of Labor Statistics, 2012).

In the rotating design, approximately one-fifth of the sample is new to the survey each quarter. Only limited data from this initial survey contact are summarized as part of the BLS published estimates. Instead the interview is used to “bound” the time frame for asking future questions on expenditures and to provide baseline data about the household.

The Interview survey is currently designed to collect detailed data on approximately 60 to 70 percent of household expenses. The detailed questions are arranged by major expenditure groupings (such as housing, transportation, clothing, and health care) and ask the respondent to “recall” purchases made for detailed items during the past three months. In order to cover an additional 20 to 25 percent of the household expenditures, the questionnaire also collects three-month average estimates of purchases of food and related items.

The first, second, and fifth interviews of a household deserve additional description. In the first interview with a new household, the survey collects demographic data for the household, inventories major durable goods within the household, and asks for only a one-month recall of expense items. Data from this initial interview are used in only limited ways in BLS expenditure estimates. Instead, data collected in this interview are primarily used for classification of the household, to help prevent duplicate expense reporting in subsequent quarters, and to minimize telescoping (a common tendency in recall surveys to report for a time period beyond the reference period). During the second and fifth interviews, the Interview survey also asks a series of questions to obtain a detailed financial profile. This profile includes income data such as salaries, unemployment compensation, alimony and child support, assets, and investments. These questions use a 12-month recall period.

Proxy reporting is currently used in the CE Interview and Diary surveys, allowing a single household representative to respond for the entire household. The accuracy of data collected from proxies depends heavily on how much the proxy respondent knows about the daily expenditures of all household members. BLS uses proxy reporting in a tradeoff for lower costs and reduced burden. Field representatives attempt to interview the “most knowledgeable” member of the household, who is more often female than male.

The BLS estimates that the average time to complete a quarterly Interview survey is approximately 60 minutes. The panel believes that many interviews are rushed, and so this time estimate may be shorter than what is needed for accurate reporting of expenditures (see p. 83 of Chapter 5, “Motivation in Interview Survey,” for more discussion on this point). The CE Interview survey does not currently offer monetary incentives to respondents.

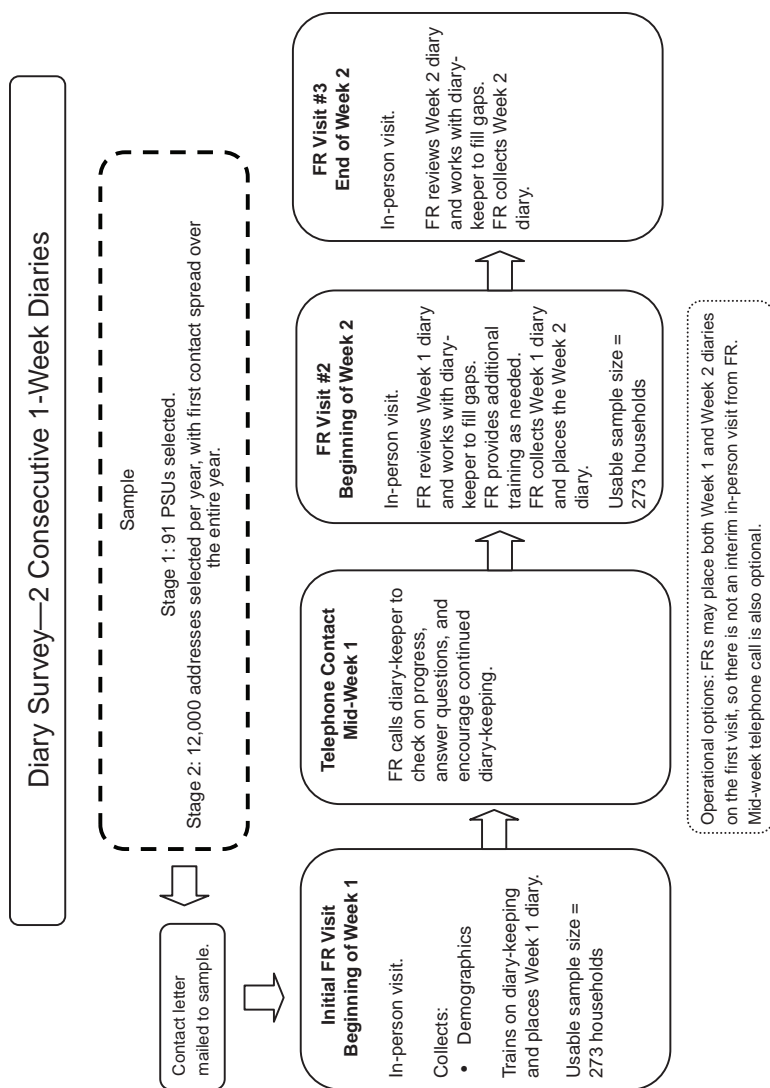


FIGURE 3-2 Process flow of the Diary survey.
NOTE: FR = field representative, PSUs = primary sampling units.
SOURCE: Panel designed flowchart based on documentation (Bureau of Labor Statistics, 2012).

Design and Implementation of the Diary Survey

Sampling Frame and Sample Size for the Diary Survey

Selection of the Diary survey sample begins with the same 91 PSUs selected for the Interview survey. Each year the Census Bureau then draws a separate sample of approximately 12,000 addresses from the augmented 2000 Census 100-Percent Detail File. The effective sample size for the Diary survey is 7,100 interviewed households, producing approximately 14,200 weekly diaries. The placement of diaries is spread equally over the 52 weeks of the year. There are approximately 273 diaries completed each week.

Implementation of the Diary Survey

Figure 3-2 displays a process flow of the Diary survey. Households selected for the Diary survey are asked to keep two sequential one-week diaries of expenditures of household members. Excluded are expenses incurred by a household member while away from home overnight. Also excluded are credit and installment plan payments made during the two-week period. The Diary survey had an overall response rate of 72 percent in 2010. (See “Comparison of Response Rates” in Chapter 5, for a discussion of how these rates are calculated.)

The Diary survey begins with a pre-survey notification letter to selected households, followed by a visit from a field representative to “place” the first week’s diary. As with the Interview survey, the Diary survey is designed for proxy reporting. A single member of the household is asked to keep the diary, recording expenditures for the household and for household members. At this initial visit, the field representative also collects data on a Household Characteristics Questionnaire about the family composition and demographics. BLS uses this information for household classification. BLS also uses this information as a way to associate Diary households with similar households from the Interview survey to analyze comparable expenditures and create integrated tabulations.

The Diary form is left with the person designated as the household respondent. It is a paper-based “self-reporting” form. The form is structured around the “day” of the purchase and by classification of whether the item was (1) food purchased away from home; (2) food purchased for consumption at home; (3) clothing; or (4) other expenditures during the week. The household respondent is asked to list any item purchased and to provide a detailed description of the item as well as its cost.

As the survey is currently designed, the field representative picks up and reviews the diary after the first week and places a second week’s form with the household diary-keeper. The field representative returns at the end of the

second week, picks up the second diary, and collects additional data on the work experience and income for the previous year of individual household members. In practice, both diaries are sometimes placed and picked up at the same time, without the intervening visit. The CE Diary survey does not currently offer monetary incentives to respondents.

COST OF THE CONSUMER EXPENDITURE SURVEYS

BLS provided the panel with an internal spreadsheet containing calendar 2010 survey costs. The cost discussions on the Interview survey and the Diary survey are based on this spreadsheet.

The total “field” cost for the CE in 2010 was \$21.2 million. Total field cost includes interviewer salaries/benefits, mileage, training, awards, and related expenses. It also includes Census staff cost in the Field Division. It does not include BLS staff costs or costs for Census employees in other divisions.

The total cost for the Interview survey in 2010 was \$17.4 million. This includes five quarters of interviews, dealing with approximately 60,000 cases. The cost per case worked in 2010 was \$283, while the cost per completed interview was \$487. Interviewing and mileage costs composed 60 percent of the total. Excluding costs associated with noninterviews, the cost for interviews completed entirely or in part via in-person interviewing was \$324 each, while those interviews completed entirely over the telephone cost \$146 each. The cost per case worked for the first quarter was approximately 14 percent higher than the average of the other four quarters. This is probably due to more screen-outs in the initial contacts and a greater use of telephone interviewing in subsequent quarters.

In fiscal 2010, the total cost for the Diary survey was approximately \$3.8 million, or approximately 17.9 percent of the total cost of both CE surveys. At the same time, the Diary survey has approximately 20 percent of the total contacts for the CE surveys.

Table 3-1 provides a summary of the design features and costs of both surveys for comparison.

TABLE 3-1 Design Features of the Consumer Expenditure Surveys

	Current Consumer Expenditure Surveys	
	Interview Survey	Diary Survey
Population	U.S. civilian noninstitutional population	U.S. civilian noninstitutional population
Information Collected	Household and personal expenditures: large and regularly occurring expenditures via 3+ month recall	Household and personal expenditures: small, frequently purchased items
Sampling Frame	2000 Census 100-Percent Detail File augmented by new construction permits	2000 Census 100-Percent Detail File augmented by new construction permits
Sample Design	Multistage area probability sample using 91 Primary Sampling Units; rotating quarterly panel design for five quarters	Multistage area probability sample using 91 Primary Sampling Units; sampled addresses are equally spread across 52 weeks
Mode and Field Protocols	Personal interview (CAPI) with decentralized telephone interviewing (CATI) allowed	Initial interview and 2 consecutive weekly diaries; weekly follow-up to retrieve diary
Household Sample Size	15,000 addresses	12,000 addresses
Average Interview Time	60 minutes (quarterly interview)	Unknown
Completed Interviews in 2010	35,843	14,599
Overall Response Rate ^a	73.4%	71.5%
Approximate Cost (2010)	\$17 million	\$3.7 million
Approximate Cost per Completed Case (2010)	\$476	\$248
Data Collection Period for Each Household	13 months	2 weeks

^aSee Chapter 5, Nonresponse section, for a fuller explanation of response rates.

SOURCES: Panel-designed table based on BLS documentation (Bureau of Labor Statistics, 2012) and internal spreadsheet provided to panel.

4

The Panel's Investigation into the Issues with the CE

In 2009, the Bureau of Labor Statistics (BLS) initiated a systematic, comprehensive study of the challenges faced by the Consumer Expenditure Surveys (CE) with the goal of redesigning the existing surveys to reduce measurement error. BLS states that the mission of this venture, known as the Gemini Project, is

to redesign the Consumer Expenditure surveys (CE) to improve data quality through a verifiable reduction in measurement error, particularly error caused by underreporting. The effort to reduce measurement error will combat further declines in response rates by balancing any expected benefits of survey design changes against any potential negative effects on response rates. Any improvements introduced as part of the Gemini Project should not increase budgetary burden, but instead, should remain budget neutral. (Bureau of Labor Statistics, 2011e, p. 1)

Since the beginning of the Gemini Project, BLS has undertaken a number of information-gathering meetings, conference sessions, forums, and workshops to aid in its mission. All of these have provided valuable information for the Panel on Redesigning the BLS Consumer Expenditure Surveys in its current task, and many of the papers presented at them are cited in this report. These events included the National Bureau of Economic Research's Conference on Improving Consumption Measurement (July 2009); Survey Redesign Panel Discussion, cosponsored by the Washington Chapter of the American Association for Public Opinion Research (DC-AAPOR) and the Washington Statistical Society (January 2010); Data Capture Technology Forum (March 2010); AAPOR Panel on Respondent Record Use (May

2010); Data User Needs Forum (June 2010); and CE Methods Workshop (December 2010). More information can be found about these events, plus copies of papers presented at them, on the BLS website (see <http://www.bls.gov/cex/geminimaterials.htm>). Additionally, BLS has conducted internal research in support of the Gemini mission and has contracted targeted research from the private sector.

The panel commends BLS on its multiyear, systematic review of the methodology used in the CE.

Building on the work of the Gemini Project, the panel investigated the opportunities and drawbacks related to the CE. As described in this chapter, their additional investigation included feedback from CE data users, panel members' reactions when they assumed the role of survey respondents, and a workshop to learn more about other large-scale household surveys. Redesign options developed by two outside groups in response to a Request for Proposal also formed an important part of the panel's investigations, and the chapter concludes with some of the main points and discussions elicited by these two options.

FEEDBACK FROM DATA USERS

The panel was diligent in reaching out to data users and trying to understand the many uses of the CE. Many of those uses are outlined in Chapter 2, including input into calculation of the Consumer Price Index (CPI), development of government programs, and as the basis for research and analysis. Several panel members are themselves regular users of the CE microdata. The panel reviewed a broad set of published research that used the CE as a source of information. As noted above, members studied the papers from the BLS 2010 Data User Needs Forum. They also attended conferences held by the National Bureau of Economic Research and held a session with microdata users at the 2011 CE Microdata Users' Conference. Finally, the panel spoke one-on-one with many users of the CE data.

The panel studied the complexities of the CPI program and how the CE supports those important indices. Considerable detail on this topic is provided in Chapter 2, in the section "CE Data Provide Critical Input for Calculating the Consumer Price Index." From their investigation, the panel made the following two conclusions.

Conclusion 4-1: The CPI is a critical program for BLS and the nation. This program requires an extensive amount of detail on expenditures, at both the geographic and product level, in order to create its various indices. The CPI is the current driver for the CE program with regard for the level of detail it collects. The CPI uses over 800 different expenditure items to create budget shares. The current CE supplies data for

many of these budget shares. However, even with the level of detail that it currently collects, the CE cannot supply all of the budget shares used by the CPI. There are other data sources from which the CPI currently generates budget shares.

Conclusion 4-2: The CPI does not utilize the panel nature of the current CE. Instead the national and regional estimates employed by the CE assume independence of households between quarters on the Interview survey, and independence between weeks on the Diary survey.

As discussed in the Chapter 2 section “The CE Provides Data Critical in Administering Government Programs,” the CE is used by a number of federal agencies to administer portions of their programs. To learn more details about this particular use of the CE, the panel held in-depth conversations with staff at these agencies. A summary of those conversations appears in Appendix C. From their investigation, the panel makes the following conclusion.

Conclusion 4-3: The administration of some federal programs depends on specific details collected from the CE. There are currently no other available sources of consistent data across years for some of these programs.

A third large group of users of the CE data are economic researchers and policy analysts from academic institutions, government agencies, and private organizations. These users work with tabular estimates produced by the BLS, and increasingly with microdata files from the CE. The panel talked with a number of these data users, researched the types of questions that their analyses addressed, and the characteristics of the CE that were important for those analyses. Many examples are provided in the Chapter 2 section, “CE Data: A Cornerstone for Policy Analysis and Economic Research.”

Much of this work is geared to understanding household behavior and how households adjust their consumption in response to changes in circumstances. These changes may be affected by personal events such as a change in income, marriage, loss of a job, retirement, the birth of a child, or the onset of a disability. Government program changes (such as tax reform, adjustments in minimum wage, or health care legislation) can also impact household behavior.

For data to be useful in this endeavor, users say it is necessary to have panel data with at least two observations. Many analysts indicate the strong advantage to having a third observation. A related issue is the length of each panel period. Data collected over a short period, such as in the cur-

rent two one-week Diary surveys, are able to answer questions related to how households respond to events that happen relatively frequently, such as receipt of monthly Social Security benefits (e.g., Stephens, 2003). However, a wide range of questions requires examining the same household both before and after a less frequent event such as a tax rebate, a job loss, or a divorce. These questions are more difficult to address with data collected over a short time period unless the sample size is rather large.

Regardless of the period over which expenditure is measured, an important complement is relevant household information over the same interval. In order to examine whether changes in household circumstances lead to changes in household consumption, these circumstances must be measured during the same period. The principal variables of interest are income, employment, retirement, disability, and marital status.

When panel data have been lacking, researchers have been able to create panels by using “synthetic cohorts.” The idea behind synthetic cohorts is that in place of following the behavior of the same individuals over time, researchers can create a panel by modeling individual household activity based on data from similar groups of households. Using these synthetic cohorts, researchers can examine the relationship between changes over time. While synthetic cohort data are more difficult to work with, they may prove useful for answering some questions. However, for a number of policy questions, synthetic cohort data do not provide a useful tool. Thus, the following conclusion is made regarding use of the CE for research and analysis purposes.

Conclusion 4-4: Economic researchers and policy analysts generally do not use CE expenditure data at the same level of detail required by the CPI. More aggregate measures of expenditures suffice for much of their work. However, many do make use of two current features of the CE microdata: an overall picture of expenditures, income, and household demographics at the individual household level; and a panel component with data collection at two or more points in time.

PANELISTS’ INSIGHT AS SURVEY RESPONDENTS

Panel members wanted to gain firsthand insight into the CE from the viewpoint of a respondent, so approximately three-quarters of panel members were interviewed by a Census field representative. Most experienced the Interview survey, one kept the Diary, and several did both. Box 4-1 provides some reactions of panel members to this experience. During the process, panel members asked their own questions of the field representatives. Thus, the interview experience for the panel was partly trying to recall and answer specific questions about their own expenditures, and partly

BOX 4-1
Reactions of Panel Members Following Their Interviews with
Field Representatives (FRs)

“The FR said I was the first respondent to EVER consult paper and electronic records extensively.”

“The FR interviewed me extremely quickly—fast talker. This seems to be the solution to respondent burden: get through as fast as possible.”

“I got a strong sense of how easy it becomes to say ‘no’ to a category simply because saying ‘yes’ so clearly leads to more trouble.”

“After the interview, the FR told me about the suspicion of government and concerns about intrusiveness that the FR regularly encounters, and it is much more intense and extreme than I had expected.”

“The diary appears to have some very significant strengths compared to quarterly recall. I did not see the immediate problems of being unable to respond to questions, as I experienced when doing the CE quarterly interview. This is a much easier task, even though at first blush it seems like keeping a diary for two weeks was going to be extraordinarily difficult.”

SOURCE: Committee on National Statistics Panel on Redesigning the Consumer Expenditure Survey (2011).

trying to understand the overall nature of the interviews as experienced by others. The field representatives made a number of comments to panel members about their “typical” respondents and what they considered normal respondent behavior. The panel believes that this entire process brought realism into their discussion of the cognitive issues and potential solutions (Committee on National Statistics Panel on Redesigning the Consumer Expenditure Survey, 2011).

HOUSEHOLD SURVEY PRODUCERS WORKSHOP:
DESCRIPTION AND INSIGHTS

Many of the problems and issues facing the CE are also faced by other large household survey programs, and the panel wanted to leverage the work done on these surveys toward solutions for the CE. In this endeavor, the panel planned and held a Household Survey Producers Workshop in June 2011 in Washington, DC. (The agenda for the workshop appears in Appendix E.) The panel would like to extend its appreciation to the presenters at this workshop for the insights they provided.

BOX 4-2**Sessions at the Household Survey Producers Workshop**

Session 1: Alternative Ways of Measuring Consumer Expenditures—International Experiences

Session 2: Designs That Add Flexibility in Data Collection Mode

Session 3: Designs That Effectively Mix Data from Multiple Surveys and/or External/Administrative Data to Produce Estimates

Session 4: Designs That Effectively Mix Global and Detail Information to Reduce Burden and Measurement Error

Session 5: Designs That Use “Event History” Methodology to Improve Recall and Reduce Measurement Error in Recall Surveys

Session 6: Diary Surveys That Effectively Utilize Technology to Facilitate Recordkeeping or Recall

NOTE: See Appendix E for the full agenda of the workshop.

The program for this workshop was built around six topics (see Box 4-2), each of which was specific enough to inform the panel’s redesign deliberations yet broad enough to be able to present different perspectives of the topics. After different presentations on a topic, one member of the panel discussed the insights that these presentations had for the CE redesign. A summary of the main points raised in the six sessions follows.

Session 1: Alternative Ways of Measuring Consumer Expenditures—International Experiences

The purpose of this session was to have representatives from other countries talk about how they collect consumer expenditures, the issues they face, and their approach to these issues. The panel was looking for differences and similarities that might inform redesign options for the U.S. CE.

Dubreuil et al. (2011) discussed how Statistics Canada redesigned its consumer expenditure survey. The new design of Canada’s Survey of Household Spending looks similar to the current CE in the United States. It uses a combination of a recall interview and 14-day diary for each selected household, with varying recall periods for different expense items. The previous Canadian design incorporated a “balance edit,” a feature that a number of users of the CE would like to see incorporated in the CE redesign. The Canadian redesign no longer includes this feature.

Horsfield (2011) discussed the UK Living Costs and Food Survey. It is

a relatively short household survey to collect regular expenditures such as rent and mortgage payments, along with retrospective information on certain large, infrequent expenditures such as those on vehicles. The program predominantly uses a Diary survey, as each individual aged 16 and over is asked to keep diary records of daily expenditures for two weeks. Children (aged 7–15) complete a simplified diary. Household members receive incentives for completing the diary: 10 pounds (\$15.68) per adult, 5 pounds (\$7.84) per child.

Borg (2011) discussed consumer expenditure surveys in Europe and the European Union's efforts to harmonize survey results. The EU countries all have their own expenditure surveys carried out under the responsibility of their national statistical offices. These surveys are generally periodic rather than annual. The primary purpose of these surveys is to produce the budget shares for the national consumer price indices, although there has been an increasing use of the information at both the national and EU levels. There remain some comparability issues among these surveys.

Session 2: Designs That Add Flexibility in Data Collection Mode

One of the primary reasons for the CE redesign is the need to update data collection strategies to create greater flexibility in the data collection mode. The Interview survey is conducted in person, with a fallback to telephone interviewing when a personal visit is not feasible. The Diary survey is dropped off and picked up in person, and the diary information is collected on paper forms. This session provided examples of how other surveys are incorporating response flexibility or newer data collection methods.

Smyth presented results from Olson, Smyth, and Wood (2011), an experiment within the ongoing Nebraska Annual Social Indicators Survey that allows the respondents to choose their mode preference. The experiment then uses the respondent's preferred mode of data collection and tests to see if this treatment makes a difference in response. In this limited experiment, they found that response rates were higher for those being surveyed in their preferred mode. They also found that Web survey response rates were lower than those with mail and phone contacts across all preference groups. However, they found that results changed within a mixed mode framework.

The Business Research & Development Innovation Survey (BRDIS), conducted by the Census Bureau for the National Science Foundation's National Center for Science and Engineering Statistics, is the nation's primary source of information on business R&D expenditures and the workforce. Hough (2011) reported that unlike its predecessor, which was sent to a single respondent within a company, the new BRDIS questionnaire is structured to allow and encourage different experts within a single business to provide responses in their areas of expertise. There are both paper and elec-

tronic versions of questionnaires. They have also developed an online toolkit to assist business respondents that includes spreadsheets, fillable PDFs, and personalized support by an account manager. Clearly, establishment surveys are different from household surveys in many ways, but there are similarities from which to extract ideas, such as multiple mode options and a toolkit for respondents. Different households and members of the same household might have a different comfort level with different collection modes. A key point from this presentation is that “one size” does not fit all respondents. The BRDIS recognizes that point up front and designs it into its methodology. Another point is the toolkit to further assist respondents.

Wine and Riccobono (2011) discussed the National Postsecondary Student Aid Study (NPSAS), a survey conducted by RTI International for the National Center for Education Statistics that mixes multiple sources of data and data collection modes with incentives to obtain and keep student respondents.

NPSAS data come from multiple sources, including institutional records, government databases, and student interviews. Detailed data on participation in student financial aid programs are extracted from institutional records. Data about family circumstances, demographics, education and work experiences, and student expectations are collected from students through a web-based multimode interview (self-administered and computer-assisted telephone interviews [CATI]). (National Center for Education Statistics, 2012)

A tailored incentive program is designed into the process to encourage early response.

Session 3: Designs That Effectively Mix Data from Multiple Surveys and/or External/Administrative Data to Produce Estimates

Some of the information collected on the CE may be available in administrative records or collected on other government surveys. This session highlighted surveys that, while collecting large quantities of information themselves, also utilize administrative records and/or combine data from other survey data collections to reduce the overall burden of the survey or to improve the overall quality of data.

Machlin (2011) described how the Medical Expenditure Panel Survey (MEPS) matches survey data to health records, combining them with information collected from household members and their medical providers. Upon completion of the household interview and obtaining permission from the household survey respondents, a sample of medical providers are contacted by telephone to obtain information that household respondents cannot accurately provide. This part of the MEPS is called the Medical

Provider Component (MPC), and information is collected on dates of visit, diagnosis and procedure codes, charges, and payments. The Pharmacy Component (PC), a subcomponent of the MPC, collects drug detail information, including National Drug Code (NDC) and medicine names, as well as date(s) prescriptions are filled, sources, and amounts of payment (Agency for Healthcare Research and Quality, 2012).

O'Brien (2011) discussed the Residential Energy Consumption Survey (RECS), which collects a multitude of information on houses, appliances, and home energy usage. It collects utility records from energy suppliers in lieu of self-reports from respondents. As part of this process, the interviewer asks household respondents to name their energy suppliers and to produce a bill from each supplier. The interviewer uses a portable scanner to scan in the bills. The Energy Information Agency then contacts the energy suppliers to obtain records for the sampled household unit for the previous year (U.S. Department of Energy, 2011)

Schenker and Parsons (2011) discussed combining data from multiple surveys to improve quality and reduce burden within the survey program of the National Center for Health Statistics (NCHS). They provided four examples:

- Combining information from the National Health Interview Survey (NHIS) and the National Nursing Home Survey to obtain more comprehensive estimates of the prevalence of chronic conditions for the elderly;
- Using information from the National Health and Nutrition Examination Survey (NHANES) to improve analyses of self-reported data on the NHIS;
- Combining information on the Behavioral Risk Factor Surveillance System with the NHIS to enhance small-area estimation; and
- Creating links between various NCHS surveys and administrative data sources such as air quality data available from the Environmental Protection Agency, death certificate data from the National Death Index, Medicare enrollment and claims data from the Centers for Medicare & Medicaid Services, and benefit history data from the Social Security Administration.

Session 4: Designs That Effectively Mix Global and Detail Information to Reduce Burden and Measurement Error

This session highlighted surveys that, while collecting large quantities of information, do so using design strategies and questionnaire modules that avoid asking every respondent for all details on each contact.

Aune (2011) discussed the Agricultural Resource Management Survey,

an expense and income survey of farming establishments conducted by the National Agricultural Statistics Service. It is an annual survey that collects detailed information related to the farming enterprise and, to a lesser extent, to the farm household. This survey has multiple modules or versions, with sample units assigned to a specific version during the selection process. Most versions are designed for personal enumeration, but one is designed for mail/Web collection. For a given expense item (such as fuel expenses), some versions will ask only the global expense item (total spent on fuel of all kinds) and others will ask a detailed breakout of that expense item (amount spent on gasoline, diesel, propane, etc.). Regardless of the version and mix of global/detail questions, all data are combined in summary estimates and contribute to the state, regional, and national estimates.

Fields (2011a) discussed the current structure of the Survey of Income and Program Participation (SIPP) and its use of both “core” and “topical” questionnaire items. The SIPP follows households for multiple waves. Core questions are asked in all waves, such as the global item “total income.” Topical questions are those that are not repeated in each wave. Topical modules are designed to gather specific information on a wide variety of subjects. Some topical modules cover items such as assets and liabilities, real estate property, and selected financial assets. In some instances, the topical questions are intermixed with core questions in the interview to make the questionnaire flow more smoothly.

Gentleman (2011) discussed two alternatives for asking questions about the entire family in the National Health Interview Study. The first alternative asks a global question “does anyone in the family. . . .” An alternative questionnaire goes through the family roster and asks individual questions for each family member. The NHIS is also used as a screening vehicle for follow-on surveys, with many detailed questions saved for those follow-on surveys. One result from their experiments on screening questions showed that respondents gave fewer “yes” answers to filters as they learned that such answers led to additional questions.

Session 5: Designs That Use “Event History” Methodology to Improve Recall and Reduce Measurement Error in Recall Surveys

This session highlighted surveys that utilize “event history” methodology to improve the quality of recalled information. The Panel Study of Income Dynamics (PSID) was the first major survey to implement “event history” methodology to improve the ability of respondents to recall information. Stafford (Beaule and Stafford, 2011) discussed the implementation of this methodology in the PSID, which has been a prototype for other surveys. They conducted a number of methodological studies as they developed this methodology.

Fields (2011b) discussed a newly redesigned SIPP that uses event history methodology, pulling from the experiences of the PSID. The SIPP staff believe they may be able to use a one-year recall period as effectively and accurately with this new methodology as the current design, which uses a four-month recall. The new design is scheduled to be operational in 2014. This presentation discussed the implementation of “event history” methodology and presented what has been learned so far with the pilot program.

Session 6: Diary Surveys That Effectively Utilize Technology to Facilitate Recordkeeping or Recall

Newer technology, such as the Web, smart phones, and portable scanners, has opened possibilities for diary surveys. This session highlighted surveys that utilize this newer technology to field innovative diary-type surveys.

The National Household Food Acquisition and Purchase Survey (FoodAPS) is a new pilot survey sponsored by the U.S. Department of Agriculture designed with an innovative approach to a food diary. Cole (2011) discussed the survey, which collects information on food sources, choices, quantities, prices, timing of acquisition, and nutrient characteristics for all at-home and away-from-home foods and beverages. It also collects household information that may influence food acquisition behaviors. The pilot uses color-coded booklets, portable scanners for receipts, regular telephone contact to encourage diary-keeping, and incentives as part of the data collection process (U.S. Department of Agriculture, 2011).

Kizakevich (2011) discussed personal diary and survey methodologies for health and environmental data collection used by RTI International. Among these examples were

- PFILES, a real-time exposure-related diary of product use and dietary consumption in the context of activity, location, and the environment. It uses Pocket PCs with headsets for use by respondents, who record survey responses and even take pictures of their environment;
- Personal Health Monitor for use by patients suffering from post-traumatic stress disorder and mild traumatic brain injuries to help clinicians monitor patients' status while observing symptoms and medication usage within the context of daily activities and environmental factors; and
- BreathEasy, an Android App, which allows a daily assessment of asthma triggers, health, and ventilation.

Bailey (2011) discussed Nielsen Life360 Program, which uses a “digital ethnography” approach to measure attitudes, preferences, and behaviors of the targeted population using mobile phone surveys, photography, Internet-based journals, video cameras, and Web surveys. The specially equipped smart phone prompts respondents to complete a short survey on an hourly basis in addition to capturing an image using the built-in camera as a picture description of their surroundings and activities in real time.

Summary of the Workshop

The panel found the workshop presentations to be highly informative and to provide important input into panel deliberations. A number of key points emerged from the prepared remarks that discussants delivered during the workshop, as amplified in subsequent discussion among panelists.

First, the international comparisons demonstrate that concerns about data quality and burden that have led to the need for a redesign of the CE are not unique to U.S. data collection efforts, although the size of and variability among the U.S. population present particular challenges. The alternate methods that the panel observed from other countries made clear that a bounding interview is not a universal method, and that it is plausible to rethink this aspect of CE administration. It was also clear to the panel that, although the methods and approaches from other nations have many strengths, they also have their own challenges, and simple wholesale adoption of those methods is unlikely to be a panacea for improving the CE.

Second, adding new modes of data collection needs to be done thoughtfully, attending carefully to whether adding new modes or providing respondents with a choice of mode increases data quality, reduces respondent burden, or reduces nonresponse sufficiently to be worth the design, operational, and analytic costs. As the Smyth presentation in session 2 illustrated, the scientific community is not yet at a point to fully understand why particular modes work for different respondents.

Third, while it is quite attractive to consider replacing or supplementing respondent-reported data with data from other sources (administrative records, data from other surveys) to reduce respondent burden and administrative costs, this is not as straightforward an enterprise as it might seem. The hurdles are notable enough—from mode and questionnaire differences, to sampling and weighting incompatibilities, privacy and confidentiality issues, linkage difficulties, increased agency efforts, data sharing difficulties, and lack of knowledge of costs—that it does not seem plausible to the panel that alternate sources could suffice in the short term. There is also considerable concern about whether external data would be consistently available over time.

Fourth, the panel was impressed by efforts in other U.S. surveys to

streamline data collection and rethink what kinds of general and specific information need to be asked of respondents. Although the immediate applications to the CE of the particular approaches described at the workshop are not entirely clear, and although much remains to be understood about the relationship between survey length and respondent burden, the panel's subsequent deliberations and proposals were influenced by such efforts.

Fifth, whether or not event history methods are the only or best way to stimulate all respondents' recall, the panel took note of the insight that emerges from studies of alternative interviewing methods. A redesigned CE needs to go as far as it can to accommodate respondents' natural ways of thinking about and recalling their expenditures, rather than asking respondents to conceive of their expenditures from the researcher's perspective. More broadly, assuming that respondents can recall purchases accurately without consulting records is problematic, and a redesigned CE needs to promote the use of records far more than current methods do.

Finally, the panel took very serious note of the opportunities for using new technologies to facilitate more direct and in-the-moment self-administered reporting of expenditures, as well as for passive measurement of expenditures. It will be important for a CE redesign to make as much use of these opportunities as feasible, and to start a new forward-thinking mode of research and production that continually assesses the changing technological landscape and prepares as much as possible for changes before they happen.

REDESIGN OPTIONS WORKSHOP: DESCRIPTION AND INSIGHTS

In order to elicit a broader perspective on possible solutions to the CE's problems, the panel sought formal input from organizations with experience in designing complex data collection methods. It is in this context that the panel initiated a Request for Proposal (RFP) and competitively awarded two subcontracts, one to Westat (project leader: David Cantor) and the second to a consortium from the University of Wisconsin–Milwaukee, University of Nebraska–Lincoln, and Abt-SRBI (Nancy Mathiowetz, project leader; Kristen Olson; and Courtney Kennedy). The Statement of Work (see Appendix D) required the subcontractors to produce a comprehensive proposal for a survey design, and/or other data acquisition process, that collects the data required for the primary uses of the current CE while addressing the following issues:

- Underreporting of expenditures
- Fundamental changes in the social environment for collection of survey data

- Fundamental changes in the retail environment (e.g., online spending, automatic payments)
- The potential availability of large amounts of expenditure data from a relatively small number of intermediaries such as credit card companies
- Declining response rates at the unit, wave, and item level

The full reports from Mathiowetz, Olson, and Kennedy (2011b) and Westat (2011c) are available online, and the panel summarizes them in this chapter. The panel would like to commend both subcontractors on the reports they submitted. Both designs were innovative and well thought out. The time frame was very short for completing this contract, and both groups met the challenge. Their work provided very valuable input into the panel's work from specific design options, use of technology, and review of relevant literature. The panel used their research and ideas extensively.

The panel hosted a Redesign Options Workshop on October 26, 2011, and an informal roundtable on October 27 to facilitate a public discussion of the two proposals and their relative merits in regard to the current CE. An agenda for the workshop is in Appendix F. Kulka (2011) discussed both reports with a focus on the cognitive issues related to the CE, and Bowie (2011) talked about issues relative to implementing major changes in a large ongoing survey. Data users also addressed the proposed redesigns from the perspective of their use of the CE data.

A number of important insights arose from the discussion of these proposals. One concerned the amount of detail that is required for the CE. A strong opinion was offered that one cannot collect that quantity of detail without a lot of measurement error. An understanding of what constitutes a tolerable measurement error must be clear, followed by a move back to collecting data at a more aggregate level for the prescribed level of quality. Another important round of discussion concerned the amount of additional research that would be needed to be ready to field a newly redesigned CE survey. If the redesign includes fairly major changes, as did the two proposals offered at the workshop, then a significant amount of targeted research will lie ahead.

Most important, these discussions led BLS senior management to modify their original charge to the panel. In this modified charge (see Appendix B), the panel is asked to view the Consumer Expenditure Survey (CE) Data Requirements (Henderson et al., 2011) as the mandatory requirements for the survey. The CPI data requirements document (Casey, 2010) was no longer a part of the mandatory requirements that the redesign would need to meet.

Redesign Proposal: Westat

Westat's proposed redesign (Westat, 2011b,c) focuses on three inter-related goals: (1) reducing respondent burden, (2) incorporating administrative and personal record information, and (3) improving self-report methodology. It calls for greater reliance on records and less reliance on respondent recall. Key features of this proposal and a link to the full report are provided in Box 4-3. The Westat proposal continues from the base of separate diary and interview surveys, but implemented differently than in the current CE. It introduces the concept of a "data repository" and a separate Administrative Record Survey to obtain certain records directly from retailers, utilities, and mortgage companies. The authors discussed their deliberations concerning access to external data:

Data obtained directly from retailers are likely to be more accurate than respondent-provided data are likely to be. Other federal surveys, such as the National Immunization Survey and the Residential Energy Consumption Survey, have employed administrative data to supplement and improve the quality of data reported by respondents. Conceivably, CE respondents could provide their loyalty card numbers to interviewers, who would then ask the retailers to provide the purchasing histories for those loyalty cards. This method would not be perfect; a consumer may sometimes forget to give a loyalty card to the cashier or may lend the card to friends. Moreover, retailers do not routinely release purchasing histories.

BOX 4-3 Key Features of the Westat Proposal

- Separate diary and interview surveys but implemented differently than the current CE.
- Multiple diary-keepers within a household.
- Data repository into which respondents can upload scanned receipts and records.
- Data electronically extracted from receipts and records, and a Web survey electronically generated to request missing information.
- Two recall interview surveys, one year apart. Variable recall periods used.
- Respondents contacted three months before recall interview and encouraged to keep and scan receipts during three-month period.
- Consent requested to obtain expenditure records directly from retailers, utilities, and mortgage companies. A separate administrative records survey to obtain those records.

NOTE: Link to full report: http://www.bls.gov/cex/redwrkshp_pap_westatrecommend.pdf and http://www.bls.gov/cex/redwrkshp_app_westatrecommend.pdf.

The BLS might explore the feasibility of obtaining purchasing history data by contacting large retailers with loyalty card programs. Expenditure data is also potentially available from utility companies, rental agents, and lenders. (Westat, 2011c, p. xiii)

For the Diary survey, each person age 14 and over in a sampled household would be asked to report expenditure data for 14 days. Having multiple respondents minimizes concerns about proxy reporting. The respondents are given a variety of reporting options. They could use the current paper diary forms, mail in their receipts and records, or report data electronically. All respondents are asked to save and then supply receipts. A key component of the redesigned Diary survey is a “data repository” into which respondents upload various types of expense records. The repository system would extract purchase data from the uploaded records/receipts and generate a Web survey that would ask the respondent to supply any remaining information that the CE program needs about those purchases. Respondents who chose to report their data electronically would be given a portable scanner. Using specially designed software, they would e-mail files of their scanned receipts and other records of purchases to the data repository.

Respondents would also be asked to download data files from various financial accounts and e-mail these files to the data repository. Respondents could opt to report their data by mailing in their receipts and financial statements, and staff would scan these receipts into the data repository. The authors discussed in their report the potential of asking respondents to supply financial records:

Consumers today commonly make purchases using modes that leave an electronic record. When an electronic record exists, respondents potentially could provide the expenditure data by retrieving information about the purchase from a database, or by printing out a record of the purchase, rather than by trying to remember the details of the purchase or by finding a receipt. For example, soon after a consumer makes a purchase using credit card, debit card, check, electronic fund transfer, or PayPal, a record of the transaction appears in a file that can be downloaded from the website of a financial institution. When a consumer makes an online purchase, the vendor typically sends a confirmation email, or provides a confirmation page, that the consumer can print out. The CE program does not currently ask respondents to provide these electronic records of expenditures. Their potential role in the CE data collection process deserves attention. These records cannot provide all of the data required by the CE program, however. They potentially offer a way to help respondents remember expenditures without a great deal of effort. (Westat, 2011c, p. 5)

Respondents would also be asked to provide consent to collect their purchasing history data directly from retailers. The authors discuss this recommendation, saying:

If CE respondents provided their loyalty card numbers, and retailers were willing to release purchasing data, the CE program would have access to objective information about the respondents' expenditures. Of course, this idea has some drawbacks. Consumers sometimes forget to provide their loyalty card to the cashier when they make a purchase. Some consumers may lend their loyalty cards to friends. Also, most retailers, including Walmart, have no loyalty card programs. (Westat, 2011c, p. 6)

A field representative would monitor the respondent's reporting activity, increasing contact and assistance to those not reporting regularly. A telephone or personal visit to the household would be scheduled after 7 and 14 days: a telephone interview for those that have been providing the information on a regular basis and a personal interview for households that have not been providing the information.

For the Interview survey, Westat proposes a change from the current data collection schedule. A new panel would enter the CE program each quarter, so that four panels would enter each year. The first wave of data collection for each panel would begin with a bounding interview, followed three months later by a recall interview. The second wave of data collection would start nine months after the recall interview. Westat (2011c, p. 53) indicated that this change is made to reduce both cost and burden: "The current design has a total of five in-person interviews per household, creating significant cost and respondent burden. Reducing this number to three in-person interviews would substantially reduce this burden and may lead to greater cooperation, fewer dropouts, and better data quality."

At the start of the second wave, the household would receive a package via U.S. mail reminding them to resume data collection activities, including keeping receipts. If the household has changed, it would receive a personal visit. Data collection would end with a recall interview three months later. Westat also proposes a change to the recall period, so that it varies by expense item (one-, three-, or 12-month recall). The proposal places a very strong emphasis on having households save receipts and use records. Respondents would also be asked to provide consent for collecting their expenditure history data directly from retailers, utilities, and mortgage companies. Respondents would be encouraged to scan receipts and records into the data repository as they receive them, rather than waiting for the field representative's return interview. As with the Diary survey, the repository would generate a Web survey based on the information still needed about the receipt/record. The field representative would monitor the number of

records/receipts coming in during the three-month period and contact by telephone households that were not turning in receipts regularly.

The redesign also includes a separate Administrative Record Survey that would be developed to obtain records directly from retailers, utilities, and mortgage companies. The emphasis on obtaining data from records rather than the respondent's memory is intended to improve data quality and reduce respondent burden.

Westat estimated that the proposed diary redesign would cost approximately 60 percent more than the current diary survey. This increase in cost is primarily attributable to having multiple diary-keepers within each household. Without a budget increase, the number of sampled households would have to be reduced accordingly, and the precision of the estimates would therefore also diminish. Westat estimates that the proposed interview redesign would cost approximately twice that of the current interview survey. The increase is attributable to the increased effort in contacting more households, an effect of reducing the number of panels. The new Administrative Record Survey contributes to cost increases reported for both surveys. The redesigned methods for the interview survey result in some increase in precision of the estimates, due to eliminating the within-household correlation across panel waves within the same year. This offset does not entirely make up for the increase in cost. The report provides a simulation of the effect on the precision of the estimates using one-, three- and 12-month reference periods (Westat, 2011c).

Redesign Proposal: Mathiowetz, Olson, and Kennedy

The Mathiowetz/Olson/Kennedy proposal (Mathiowetz, Olson, and Kennedy, 2011a,b) recommends a single integrated sample design, with two components: (1) a cross-sectional one-month diary, and (2) a panel component for which a household would complete the one-month diary for three different waves within the year. The proposed design makes extensive use of tablet computers, receipt scanners, and flexible memory "triggers." Box 4-4 provides key elements of this proposal and a link to the full report. The design provides for active monitoring of the diary-keeping activities of household members, with interventions when this activity appears inadequate. Their design minimizes the reliance on retrospective recall, eliminates the need to combine data from two distinct surveys, and provides an important panel component within the data structure.

In discussing the advantages to their proposed design, the authors stated

Our design addresses the issue of underreporting by minimizing reliance on retrospective reporting, promoting "real time" recording of all ex-

BOX 4-4**Key Features of the Mathiowetz/Olson/Kennedy Proposal**

- Diary only
- Cross-sectional sample households keep diary for one month. Panel component keeps one-month diary three times during the year.
- Multiple diary-keepers within the household
- Use of tablet PCs with Internet connection to permit real time uploading of data
- Ongoing monitoring of uploaded data with feedback to household
- Use of memory triggers encouraged

NOTE: Link to full report: http://www.bls.gov/cex/redwrkshp_pap_abtsrbirecommend.pdf.

penditures 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 each member of the CU can log on to his or her own personal diary to record expenditures. 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. (Mathiowetz, Olson, Kennedy, 2011b, p. 11)

Each adult (age 16 and older) member of the selected household would be asked to keep a 30-day diary, reporting expenditures “real-time” during that period. Younger children (aged 7–15) would be asked to keep a “mini diary” for that same time period. During an initial personal visit to the selected household, the field representative would collect demographic and socioeconomic data, including asking some global questions related to certain expenditures and annual income. The field representative would probe about regular monthly payments for housing and utilities, and any automatic payment schedules. The 30-day diary process would be explained with appropriate training on use of the diary tools. The authors provide their rationale for multiple diary-keepers within the household.

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 means for improving data quality (Bound, Brown and Mathiowetz 2001). 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). (Mathiowetz, Olson, and Kennedy, 2011b, p. 11)

During the 30-day diary period, household members would be asked to keep receipts and record expenditures on a real-time basis using one or more of the diary tools provided, with a computer tablet with Internet connection as the primary recording tool. The tablet would be available for use by all household members. It would feature an instrument that minimizes burden and facilitates consistency in reporting of required details. An attached scanner and bar code reader would facilitate data capture of products and receipts. The proposal also recommends the use of a personalized e-mail account to forward receipts and electronic records. The field representative would conduct a “wrap-up” interview and data review at the end of the diary period, with retrospective questions asked as needed to fill gaps in the diary-keeping.

Mathiowetz/Olson/Kennedy encourage the adoption of multiple portable means of capturing triggers that help household members remember a purchase so that it can be recorded later. These include the respondent’s own smart phone to record pictures, voice recordings, and notes. A small, simple pocket diary could also be used as a memory trigger.

A panel component of the design is recommended to better support micro-level analysis for the entire year. It is formed by a subset of the overall sample that is asked to complete a 30-day diary for months 1, 7, and 13. The authors deliberated on the best length for the diary reporting interval and the panel, stating that:

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. (Mathiowetz, Olson, and Kennedy, 2011b, p. 15)

This proposal recommends that BLS continue to look at sources of administrative data for benchmarking and microlevel use. Mathiowetz/Olson/Kennedy discuss a number of existing data sources, including three federal surveys that might be used to benchmark CE data. The authors also discuss nonfederal sources of data but do not incorporate a specific recommendation for their use into the current proposal. They state that they “were initially optimistic about micro-level integration of non-federal administrative data sources with CE data. However, the current state of knowledge about these 16 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” (Mathiowetz, Olson, and Kennedy, 2011b, p. 16).

Summary of the Two Proposals

While the panel does not recommend implementing either of these two designs wholesale, the designs embody important insights that became central to its deliberations, and aspects of each design are incorporated into one or all of the panel’s three proposed designs presented in Chapter 6. Both proposals place renewed emphasis on the use of survey personnel to provide help, consultation, and monitoring of respondents’ efforts, and the panel’s thinking was clearly inspired by this model.

The most notable adoption from the Mathiowetz/Olson/Kennedy proposal is a focus on supported self-administration and the use of a tablet data collection interface. These concepts are a central feature in all three of the panel’s prototypes described in Chapter 6. One prototype, *Design A, Detailed Expenditures Through Self-Administration*, follows much of the Mathiowetz/Olson/Kennedy proposal, as does the diary component of *Design C, Dividing Tasks Among Multiple Integrated Samples*. The panel’s proposed designs were inspired, in different ways, by the Westat proposal’s strong focus on encouraging the use of records. *Design C, Dividing Tasks Among Multiple Integrated Samples* follows the Westat design that encourages respondents to keep receipts and record expenditures throughout the quarter prior to a visit by the field representative. The Westat data repository proposal was viewed as desirable in the future but less practical in the nearer term.

5

Why Redesign the CE?

Today, all household surveys, including the Consumer Expenditure Surveys (CE), face well-known challenges. These challenges include maintaining adequate response from increasingly busy and reluctant respondents. In addition, more and more households are non-English speaking, and a growing number of higher-income households have controlled-access residences. Call screening and cell-phone-only households have made telephone contacts on surveys more difficult. Today's household surveys face confidentiality and privacy concerns, a public growing more suspicious of its government, and competition from an increasing number of private as well as government surveys vying for the public's attention (Groves, 2006; Groves and Couper, 1998).

In the midst of these challenges for household surveys, the CE surveys stand out as particularly long and arduous. In the Interview survey, recall of many types of expenditures is likely to be imperfect. A typical respondent lacks records or at least the motivation to use them in answering the CE questions. The level of detail that is required in describing each purchase is daunting. In the Diary survey, respondents are asked to record the details of many small purchases in a complicated booklet. These demands can easily result in limited compliance and the omission of expenditures.

Further exacerbating the problem, the CE faces the additional challenge that consumer spending has changed dramatically over the past 30 years, and it continues to change (Fox and Sethuraman, 2006; Kaufman, 2007; Sampson, 2008). When the CE was designed in the 1970s, there was no online shopping or options for electronic banking and bill paying. Over that

time, shopping patterns have shifted from individual purchases at a variety of neighborhood stores to collective purchasing at “big box” stores such as Walmart, Target, and Costco that sell everything from meat to shirts, furniture, and motor oil under one roof. The CE surveys are cognitively designed to collect spending information based on the 1970s world of purchasing behaviors, and today’s consumers are unlikely to relate to that.

Underreporting of expenditures is a long-standing problem with the CE as evidenced by a growing deviation from other data sources and by the results of several studies. This underreporting appears to differ sharply across commodities, raising the possibility of differential biases in the Consumer Price Index (CPI) and the picture of the composition of household spending. This is the biggest concern with the CE program. The Panel on Redesigning the BLS Consumer Expenditure Surveys believes that there are a number of issues with the current design and implementation of the CE, and that collectively these problems lead to the underreporting of expenditures. This chapter documents this underreporting and then discusses the issues and concerns that the panel identified in its study of the CE.

With that said, the panel understands that no survey is perfect. In fact all surveys are compromises between the need for specific data, the quality with which those data can be collected, and the burden and costs required to do so. The CE is no exception. It is the panel’s expectation that by examining the issues with the current CE along with some alternative designs, a new and better balance can be found between data requirements, data quality, and burden.

EVIDENCE OF UNDERREPORTING IN THE CE

In many federal surveys, one can assess the quality of data by comparisons with other sources of information. One of the difficulties in evaluating the quality of CE data is that there is no “gold standard” with which to compare the estimates. However, several sources provide insight into data quality. The National Research Council, in its review of the conceptual and statistical issues with the CPI, expressed concern about potential bias in the expenditure estimates from the CE. That report recommended comparison of the CE estimates with those from the Bureau of Economic Analysis’s Personal Consumption Expenditures (PCE):

The panel’s foremost concern is with the extent of bias in the CEX [Consumer Expenditure Surveys] which, in turn affects the accuracy of CPI expenditure category budget shares. A starting point for evaluating household expenditure allocations estimated by the CEX is to compare them against budget shares generated by other sources. The Bureau of Economic Analysis (BEA) produces the most obvious alternative, the per-capita and product accounts (NIPA). (National Research Council, 2002, p. 253)

Comparisons Between the CE and PCE

Compatibility

A long literature has focused on the discrepancy between the CE and PCE data from the National Income and Product Accounts (NIPA) (Attanasio, Battistin, and Leicester, 2006; Branch, 1994; Garner, McClelland, and Passero, 2009; Garner et al., 2006; Gieseman, 1987; Meyer and Sullivan, 2011; Slesnick, 1992). However, in comparing the CE to the PCE data, it is important to recognize conceptual incompatibilities between these data sources. Slesnick (1992, p. 22), when comparing CE and PCE data from 1960 through 1989, concluded that “approximately one-half of the difference between aggregate expenditures reported in the CEX [CE] surveys and the NIPA can be accounted for through definitional differences.” Similarly, the General Accounting Office (1996, p. 15), now the U.S. Government Accountability Office, in a summary of a Bureau of Economic Analysis comparison of the differences in 1992, reported that “more than half was traceable to definitional differences.”

Thus, a key conceptual difference between the CE and PCE is “what is measured.” The CE measures out-of-pocket spending by households, while the PCE definition is wider, including purchases made on behalf of households by institutions. The CE is not intended to capture purchases by households abroad such as those on military bases, whereas the PCE includes these purchases. These differences are important and growing over time. Imputations including those for owner-occupied housing and financial services, but excluding purchases by nonprofit institutions serving households and employer contributions for group health insurance, now account for over 10 percent of the PCE. In-kind social benefits account for nearly another 10 percent. Employer contributions for group health insurance and workers’ compensation account for over 6 percent, while life insurance and pension fund expenses and final consumption expenditures of nonprofits represent almost 4 percent. McCully (2011) reported that in 2009 nearly 30 percent of the PCE was out-of-scope for the CE, up from just over 7 percent in 1959.

Another important conceptual difference between the CE and PCE is the underlying data and how the estimates are constructed. Chapter 3 of this report describes the CE surveys in some detail. In comparison, the PCE aggregates come from data on the production of goods and services, rather than consumption or expenditures by households. The PCE depends on multiple sources, primarily from business records reported on the economic censuses and other Census Bureau surveys. The PCE numbers are the product of substantial estimation and imputation processes that have their own error profiles. Estimates from these business surveys are adjusted using input-output tables to add imports and subtract sales that do not go

to domestic households. These totals are then balanced to control totals for income earned, retail sales, and other benchmark data (Bureau of Economic Analysis, 2010, 2011a,b).

One indicator of the potential error in the PCE is the magnitude of the revisions that are made from time to time (Gieseman, 1987; Slesnick, 1992). A recent example is the 2009 PCE revisions, which substantially revised past estimates of several categories. *Food at home*, one of the largest categories, decreased by over 5 percent after the 2009 revision.¹

Some authors have argued that despite the incompatibilities between the CE and PCE, the differences between the series should be expected to be relatively constant (Attanasio et al., 2006). While a plausible conclusion, a gradual widening of the difference between the sources could still be expected given their growing incompatibility, as reported in McCully (2011) and Moran and McCully (2001).

Comparisons

Gieseman (1987) conducted one of the first evaluations of the current CE, comparing the CE to the PCE for 1980–1984.² He found that the CE reports were close to the PCE for *rent, fuel and utilities, telephone services, furniture, transportation, and personal care services*. On the other hand, substantially lower reporting in the CE for *food, household furnishings, alcohol, tobacco, clothing, and entertainment* was apparent back in 1980–1984.

The current patterns have strong similarities to those from 30 years ago. Garner et al. (2006) reported a long historical series of comparisons for the integrated data that begins in 1984 and goes up through 2002. Some categories compare well. *Rent, utilities, and fuels and related items* are reported at high and stable levels relative to the PCE. *Telephone services, vehicle purchases, and gasoline and motor oil* are reported at high levels (compared to the PCE) but have declined somewhat over time. *Food at home* relative to the PCE is about 0.70, but has remained stable over time. The many remaining categories of expenditures are reported at low levels relative to the PCE, though some small categories such as *footwear and vehicle rentals* show relative increases.

¹The 2008 value for food at home was \$741,189 (in millions of dollars) prior to revision and \$669,441 after, but the new definition excludes pet food. A comparable pre-revision number excluding pet food is \$707,553. The drop from \$707,553 to \$669,441 is 5.4 percent. Appreciation is given to Clinton McCully (BEA) for clarifying this revision.

²Comparisons of consumer expenditure survey data to national income account data go back at least to Houthakker and Taylor (1970). The issues were also addressed in a long series of articles comparing the CPI to the PCE deflators by Bunn and Triplett (1983) and Triplett and Merchant (1973).

Garner et al. (2006) ultimately argued that this comparison should focus on expenditure categories whose definitions are the most comparable between the CE and PCE, noting “a more detailed description of the categories of items from the CE and the PCE is utilized than was used when the historical comparison methodology was developed. Consequently, more comparable product categories are constructed and are included in the final aggregates and ratios used in the new comparison of the two sets of estimates” (Garner et al., 2006, p. 22). The new series provides comparisons every five years from 1992 to 2002 (Garner et al., 2006), and were updated and extended annually through 2007 in Garner, McClelland, and Passero (2009).

When using comparable categories and when the PCE aggregates are adjusted to reflect differences in population coverage between the two sources, the ratio of total expenditures on the CE to PCE is fairly high but still decreases over time. The ratio for 1992 and 1997 was 0.88, while in 2002 it was 0.84 and by 2007 had fallen to 0.81 (Garner, McClelland, and Passero, 2009). Figure 5-1 shows the time pattern for the ratio of CE to PCE spending for comparable categories over 2003–2009. The above discussion highlights that it is easy to overstate the discrepancy between the CE and the PCE by comparing all categories, rather than restricting the comparison to categories with comparable definitions (Passero, 2011).

Separate Comparison of the Interview Survey Estimates and the Diary Survey Estimates with the PCE

It is also important to look at comparability with the PCE of estimates from the Interview survey and Diary survey separately. Gieseman (1987) reported separate comparisons of the Interview survey and Diary survey estimates to PCE estimates for *food* because these were the only estimates available from both surveys.³ He found that Interview *food at home* exceeded Diary *food at home* by 10 to 20 percentage points, but was still below the PCE. For what was then a much smaller category, *food away from home*, the Diary aggregate exceeded the Interview aggregate by about 20 percentage points. Again, the CE numbers were considerably lower than the PCE ones.

It is not surprising that the Interview and Diary surveys yield different estimates, given the different approaches to data collection, including a

³In these early years, BLS published separate tables for Interview and Diary data. In recent years, tables have been published with only integrated data. Consequently, subsequent comparisons of CE to PCE almost exclusively rely on the integrated data that combine Interview survey and Diary survey data. In cases where the expenditure category is available in both surveys, the BLS selects the source for the integrated data that is viewed as most reliable. See Creech and Steinberg (2011) and Steinberg et al. (2010).

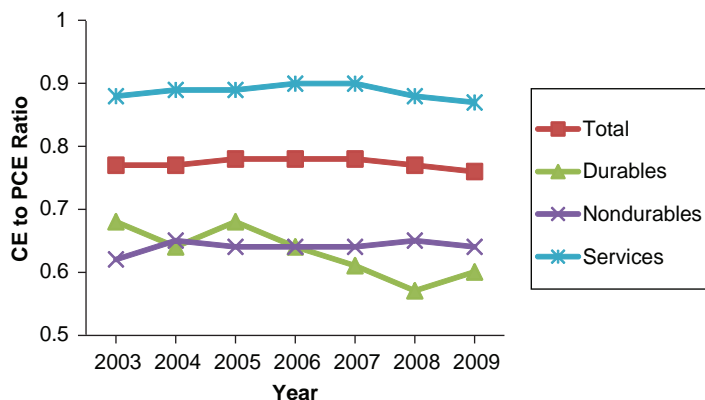


FIGURE 5-1 Coverage of comparable spending between the CE and PCE.

NOTE: CE = Consumer Expenditure Surveys, PCE = Personal Consumption Expenditures.

SOURCE: Passero (2011).

different form of interaction with the respondent household. These differences provide the likelihood of differences in estimates between the two surveys as currently configured, as discussed in more detail later in this chapter.

Bee, Meyer, and Sullivan (2012) looked further at comparing the estimates from both surveys separately to the PCE. The authors examined estimates for 46 expenditure categories for the period 1986–2010 that are comparable to the PCE for one or both of the CE surveys. Table 5-1 shows the 10 largest expenditure categories for which these separate comparisons can be made, showing ratios of the CE to PCE for these categories. Among these categories, six (*imputed rent on owner-occupied nonfarm housing, rent and utilities, food at home, gasoline and other energy goods, communication, and new motor vehicles*) are reported on the CE Interview survey at a high rate (relative to the PCE) and have been roughly constant over time. These six are all among the eight overall largest expenditure categories. In 2010, the ratio of CE to PCE exceeded 0.94 for *imputed rent, rent and utilities, and new motor vehicles*. It exceeded 0.80 for *food at home* and *communication* and is just below this number for *gasoline and other energy goods*. In contrast, no large category of expenditures was reported at a high rate (relative to the PCE) in the Diary survey that was also higher than the equivalent rate calculated from the Interview survey. Reporting of *rent and utilities* is about 15 percentage points higher in the Interview survey than the Diary survey. *Food at home* is about 20 percentage points higher in the

TABLE 5-1 CE/PCE Comparisons for the 10 Largest Comparable Categories, 2010

PCE Category	PCE (\$ millions)	Ratios	
		Diary to PCE	Interview to PCE
Imputed rental of owner-occupied nonfarm housing	1,203,053		1.065
Rent and utilities	668,759	0.797	0.946
Food at home	659,382	0.656	0.862
Food away from home	545,579	0.519	0.506
Gasoline and other energy goods	354,117	0.725	0.779
Clothing	256,672	0.487	0.317
Communication	223,385	0.686	0.800
New motor vehicles	178,464		0.961
Furniture and furnishings	140,960	0.433	0.439
Alcoholic beverages purchased for off-premises consumption	106,649	0.253	0.220

NOTE: CE = Consumer Expenditure Surveys, PCE = Personal Consumption Expenditures.
SOURCE: Bee, Meyer, and Sullivan (2012).

Interview survey.⁴ *Gasoline and other energy goods* are about 5 percentage points higher in the Interview survey and *communication* is about 10 percentage points higher. The 2010 ratios for *food away from home* and *furniture and furnishings* are close to a half for both the Interview and Diary surveys. For *clothing* and *alcohol*, the Diary survey ratios are below 0.50, but the Interview survey ratios are even below those for the Diary survey.

The panel next looked at smaller expenditure categories that are comparable between the PCE and the CE. Of the 36 such categories, only six in the Interview and five in the Diary have a ratio of at least 0.80 in 2010. In the Diary survey *household cleaning products* and *cable and satellite television and radio services* were reported with a high rate (comparable to the PCE). *Household cleaning products* had a ratio (relative to the PCE) of 1.15 in 2010 in the Diary survey; the ratio has not declined appreciably in the past 20 years. The largest of these categories reported with a high rate (comparable to the PCE) in the Interview survey were *motor vehicle accessories and parts*, *household maintenance*, and *cable and satellite television and radio services*. The remaining categories were reported at low

⁴There is some disagreement about how to interpret the fact that food at home from the CE Interview survey compares more favorably to PCE numbers than does food at home from the CE Diary survey. Some have argued that the CE Interview survey numbers may include nonfood items purchased at a grocery store. Battistin (2003) argued that the higher reporting of food at home for the recall questions in the Interview component is due to overreporting, but Browning, Crossley, and Weber (2003) stated that this is an open question.

rates (compared to the PCE) in both surveys with ratios below one-half. These include *glassware, tableware, and household utensils* and *sporting equipment*. *Gambling and alcohol* had especially low ratios, below 0.20 and 0.33, respectively, in both surveys in most years.

Summary of Comparisons with the PCE

The overall pattern indicates that the estimates for larger items from the CE are closer to their comparable estimates from the PCE. The current Interview survey estimates these larger items more closely to the PCE than does the current Diary survey. For the 36 smaller categories, neither the Interview survey nor the Diary survey consistently produces estimates that have a high ratio compared to the PCE. The categories of expenditures that had a low rate (compared to the PCE) tended to be those that involve many small and irregular purchases, categories of goods for specific family members (*clothing*), and categories for which individuals might want to underestimate purchases (*alcohol, tobacco*). Large salient purchases (like *automobiles*), and regular purchases (like *rent and utilities*) for which the Interview survey was originally designed, seem to be well reported. These patterns have been largely evident since the 1980s or even earlier. However, over the past three decades, there has been a slow decline in the level of reporting of many of the mostly smaller categories of expenditures in both the Interview survey and the Diary survey.

Similar results are reported from Canada. Statistics Canada's consumption survey was redesigned with both a recall survey and diary, with partial implementation in 2009. The level of expenditures from the diary was found to be 14 percent less than the recall interview for less frequent expenses and 9 percent less for frequent expenditures. Incomplete diaries contributed to the underestimation, given that 20 percent of diary days were "nonresponded" days (Dubreuil et al., 2011).

The panel reiterates that there are many differences between the CE and the PCE, and it does not consider the PCE to be truth. Nevertheless, the most extensive benchmarking of the CE is to the PCE, so these results are informative. Furthermore, when separate comparisons of the Interview survey and the Diary survey to the PCE are available, the comparisons provide an indication of the possible degree of relative underreporting in the two surveys.

Comparisons Between the CE and Other Sources

There have been comparisons of the CE to a number of other sources. Most are summarized on the BLS Comparisons Web page.⁵ These compari-

⁵See <http://www.bls.gov/cex/cecomparison.htm>.

sons include, but are not limited to: *utilities* compared to the Residential Energy Consumption Survey (RECS); *food at home* compared to trade publications *Supermarket Business* and *Progressive Grocer*; and *health* expenditures compared to the National Health Expenditure Accounts (NHEA) and the Medical Expenditure Panel Survey (MEPS). Some of the findings are presented below.

The CE's estimates for *utilities* are compared to those generated by the RECS. The populations of households from these two surveys are not identical, but fairly consistent. The RECS collects most information on utilities directly from utility companies after obtaining permission from the sampled households. Between 2001 and 2005, the CE estimates of total expenditures for residential energy were between 7 and 9 percent higher than from the RECS. When the energy source was broken down, the CE was higher for electricity and natural gas, while lower for the smaller category of fuel oil and LP gas.

In 2007, the CE's estimate for total health expenditures was 67 percent of the total out-of-pocket health expenditures estimated from the NHEA. The NHEA is based on a broader population definition than is the CE, and the differences between its estimates and the CE may be affected by the population differences plus the concepts, context, and scope of data collection. When compared to the MEPS, the CE estimates were lower for total health expenditures, with comparison ratios similar as those of the NHEA.

Comparisons were made between total *food at home* from the CE with grocery trade association data from *Supermarket Business* and *Progressive Grocer*. During the 1990s, the CE estimate was consistently between 10 percent and 20 percent higher than the trade association data.

Summary of Comparisons with Other Sources

The panel was not charged with evaluating the error structure of the PCE or other relevant sources of administrative data. However, the above analysis provides important background for making decisions about the CE redesign. It indicates that the concerns about underreporting of expenditures in both the CE Diary and CE Interview surveys are warranted. For many uses of the CE, any underreporting is problematic. However, for the use in calculating CPI budget shares, the differential underreporting that is strongly indicated by these results, and discussed in more detail on p. 105 of Chapter 5, "Disproportionate Nonresponse," is especially problematic. In principle, an attentive, motivated respondent could report a particular expenditure—a pound of tomatoes for a certain price—concurrently with better accuracy than in a recall survey. This potential is not evident from the estimates of aggregate spending obtained from the current designs of the CE Interview and Diary surveys. The above analysis indicates that there

are issues with both the CE Diary and CE Interview surveys, leading to the need for them to be assessed and redesigned. As a result, the panel reached this conclusion:

Conclusion 5-1: Underreporting of expenditures is a major quality problem with the current CE, both for the Diary survey and the Interview survey. Small and irregular purchases, categories of goods for specific family members, and items that may be considered socially undesirable (alcohol and tobacco) appear to suffer from a greater percentage of underreporting than do larger and more regular purchases. The Interview survey, originally designed for these larger categories, appears to suffer less from underreporting than does the Diary survey in the current design of these surveys.

MEASUREMENT DIFFERENCES BETWEEN THE INTERVIEW AND DIARY

Before examining potential sources of response errors in the Interview survey and Diary survey separately, this section considers whether these two independent surveys, as currently designed, are inherently comparable in the information that each collects. In the section above, the panel raised its concern about basic comparability of expenditure categories when comparing to the PCE. Here, the report explores another aspect of comparability.

It is important to remember the purposes for which the two surveys were originally designed. The Diary was designed to gather information on the myriad of frequent, small purchases made on a daily basis. These items include food for home consumption and other grocery items such as household cleaning and paper products. The Diary also is the source of expenditures for some clothing purchases, small appliances, and relatively inexpensive household furnishings, as well as the source of estimates on food away from home. The Interview, on the other hand, was designed to produce estimates for regular monthly expenditures like rent and utilities. It was designed to capture major expenditures, including those for large appliances, vehicles, major auto repair, furniture, and more expensive clothing items. Given the very different purposes of the two surveys, it is not surprising that they have entirely different designs and, hence, different problems.

Differences in Questions, Context, and Mode

A broad base of literature in survey research has identified many factors that can independently affect the accuracy of answers to survey questions. Some of the most important include the following:

- Different question wording is likely to produce different responses (Groves et al., 2004).
- The context in which questions are asked—for example, the purpose of the survey as it is explained to the respondent and the order in which questions are asked—influences what respondents will report (Tourangeau and Smith, 1996; Tourangeau, Rips, and Rasinski, 2000).
- Survey mode influences answers. For example, the literature demonstrates that in-person interviews are more likely to produce socially desirable answers that put the respondent in a more favorable light (Dillman, Smyth, and Christian, 2009).
- For self-administered diaries, the visual layout and design can have a dramatic effect on respondent answers (Christian and Dillman, 2004; Tourangeau, Couper, and Conrad, 2007).

These influences are realized as respondents go through the well-established cognitive process of *comprehending* the question and concluding what they are being requested to do, *retrieving relevant information* for formulating an answer, *deciding* which information is appropriate and adequate, and *reporting*. It is well documented that errors may occur at each of these stages (Tourangeau, Rips, and Rasinski, 2000).

As noted earlier, Bee, Meyer, and Sullivan (2012) found that the level of reported expenditures for certain purchases are consistently different in the Interview survey and the Diary survey. Although the Interview survey generally yields larger expenditures, these differences are not consistently in the same direction. For example, *food purchased away from home*, payments for *clothing and shoes*, and purchases for *alcoholic beverages* are greater from the Diary. Expenditures for *rent and utilities*, *food at home*, and *gasoline and other energy goods* are larger from the Interview. Some argue that larger is simply more accurate, but that may not be the case. The panel has not said that either approach or type of question is inherently better or worse. However, it is appropriate to illuminate these differences more closely.

Different questions are asked in the Interview and the Diary surveys, and these different questions are also asked in different survey contexts. To illustrate this, consider the category of *food and drink at home* to see how each survey collects this information. This is one of the categories for which the Diary was designed.

The Interview survey asks the following questions:

- What has been your or your household usual WEEKLY expense for grocery shopping? (Include grocery home-delivery service fees and drinking water delivery fees.)

- About how much of this amount was for nonfood items, such as paper products, detergents, home cleaning supplies, pet foods, and alcoholic beverages?
- Other than your regular grocery shopping already reported, have you or any members of your household purchased any food or non-alcoholic beverages from places such as grocery stores, convenience stores, specialty stores, home delivery, or farmer's markets? What was your usual WEEKLY expense at these places?
- What has been your or your household's usually MONTHLY expense for alcohol, including beer and wine to be served at home?

Thus, the Interview survey asks the respondent to estimate "usual" weekly expenditure (at grocery stores and home delivery) and to estimate a second "weekly" amount for nonfood items that is included in the first estimate. The respondent is then asked to estimate a third "weekly" expenditure for *food at home* purchased at all other places apart from grocery stores. Finally, the respondent is asked to make a fourth estimate, this time for the "monthly" purchase of alcoholic beverages consumed at home. In the Interview survey, the questionnaire does not use the term "food or drink for home consumption" but instead talks about "weekly grocery shopping" with no mention of home consumption.

In contrast, the Diary is introduced to the respondent as wanting specific expenses as the respondent makes them. Thus, the respondent is asked to individually record each purchase that fits under the category of *food and drinks for home consumption*. The emphasis here is on specific products and their detailed characteristics, and whether it is purchased for someone not "on your list." Alcohol is to be included, but the cost for alcohol is also recorded separately. The respondent is not asked to estimate any "weekly" or "monthly" amounts.

In addition, certain of these expenditures may be viewed as socially undesirable (e.g., alcohol use). An extensive literature has shown that questions about socially undesirable behaviors tend to be underreported in the presence of an interviewer and that more accurate data may be obtained in self-administered modes (Kreuter et al., 2011; Tourangeau and Smith, 1996).

Obviously, not all questions about expenditures in the CE are about socially undesirable behaviors, although questions of finance (in particular income) tend to be seen as sensitive by U.S. respondents. However, the presence or absence of an interviewer is another clear difference between the Diary and Interview collections. It is unclear what percentage of CE questions is likely to benefit from self-administration, as opposed to benefiting from having an interviewer available to clarify confusing concepts and provide motivation. The designs proposed in Chapter 6 attempt to address the un-

resolved questions about the benefits of interviewer- and self-administration in different ways. Ultimately, the panel agrees that more research will be needed to fully determine when and for which respondents it will be possible to gain the benefits of increased disclosure in self-administration while also gaining the benefits of interviewer support.

In sum, the CE Interview and Diary present quite different questions and settings so that different answers are to be expected. The Diary uses an itemization process as expenditures are occurring (using instructions that may or may not be understood). In the example of *food at home*, the Interview survey uses a “global question” quick-response format that involves addition and subtraction of items to form totals. The Interview survey includes an interviewer whose presence may act as a cheerleader or motivator, but in some cases may reduce disclosure of sensitive information or in some ways license the inference that estimation and satisficing⁶ are sufficient in order to maintain the speed of the interview. It is not surprising that different amounts are reported in the Interview and Diary in this situation, and that these differences are not always in the same direction.

Error Structure

It was beyond the resources of the panel to examine fully the error structure of the current Interview and Diary surveys. However, as the panel went through the process of considering design alternatives for the CE, there was considerable discussion about the error structure of the current surveys.

One issue of discussion was whether the different collection modes in the CE were more or less likely to produce an asymmetric error structure, and if such were the case, whether that type of structure could contribute to the differential underreporting observed between the Interview and Diary surveys. A collection process with a pronounced asymmetric error structure might be more likely to create an observable bias in the estimates.

The current Diary survey asks the diary-keeper to enter expenditures concurrently using records or short-term recall. Errors of omission—forgetting to record a purchase—are the types of errors most likely to occur. If the diary-keeper does not enter expense items on a daily basis but waits until the end of the recording period, there are likely to be more expense items left off of the diary form (more errors of omission). It is possible that the diary-keeper may recall that a purchase was made but then over- or underestimate the amount spent. This latter type of recall error might be more

⁶In this context, “satisficing” is responding to a survey question with an answer that is “good enough” to move forward to the next question, without necessarily being an accurate or complete response.

symmetrical in its structure. In general, most panel members concluded that the current Diary survey had an error structure with asymmetrical properties. This led the panel to look at ways to minimize errors of omission.

The current Interview survey collects expenses retrospectively over a three-month period. The respondent is asked to use records to report expenses, but in reality the survey depends heavily on the respondent's recall of making specific purchases over a three-month period. In this scenario, errors of omission—failure to recall a purchase or other expense—are likely to be a common type of error. As discussed above, this type of error is likely to have an asymmetrical structure. However, this problem may be less prevalent if respondents estimate expenditures during the recall period rather than trying to reconstruct all purchases as discussed below. Another common type of error is when the respondent recalls that a purchase was made, but he or she has trouble recalling the exact amount of the purchase. This type of error may have a more symmetrical structure if a respondent is as likely to over- or underestimate the amount. Moreover, the current Interview survey uses a bounding interview as its first wave. A major purpose of this bounding interview is to control for asymmetric telescoping errors (erroneously reporting an expenditure that occurred before the reference period) in recall. Some panel members hypothesized that the structure is more likely to be symmetrical in nature, not necessarily subject to bias, although the data are not available to test that conjecture.

The current Interview survey also features another type of question whose error structure may be quite different. For certain frequently purchased items (such as *gasoline* or *food at home*), the respondent is not asked to recall all purchases over the three months. Instead he or she is asked to “estimate” the usual amount the household spent on the item per month or per week. This type of question is illustrated earlier in this section. The assumption in this type of question is that the household is likely to make many such purchases and that a systematic recall of individual purchases over three months would be very problematic. So the respondent makes an “estimate” of how much the household typically spends on the item. The panel discussed the possible error structure of these types of questions. Some panel members hypothesized that the structure is symmetrical in nature, not necessarily subject to bias. Other panel members held that the panel did not have sufficient information on the structure to draw a conclusion.

Summary of Relative Error in Reporting

Back to the example, which survey more accurately collects the *food and drink consumed at home* data required by the CE? It depends. If the conclusion is based on the fact that aggregating the Interview estimates

more closely approximates the PCE total, the current Interview survey may be more accurate for this expense item. On the other hand, if based on the hypothesized accuracy of a diary response in its report of a particular expenditure, one might conclude that the Diary may be more accurate. Drawing a conclusion as to whether the Interview or Diary collects more accurate data is not possible with the data at hand. The questions are different, the context is different, and the question order is different. The field representative has greater presence in one mode. In addition, each data collection mode is subject to different causes of inaccurate reporting. The Interview relies on estimates often given with little prior thought to the exact question that is going to be asked. At the same time, Diary responses rely on “near daily” compliance to record every single expenditure, both large and small. The modes are subject to different visual layout effects. The panel knows of no research on consumer expenditures that controls for these factors while asking the same consumer expenditure questions. The panel has made the following conclusion:

Conclusion 5-2: Differences exist between the *current* Interview and Diary reports of expenditures. Differences in questions, context, and mode are likely to contribute to these differences. The error structures for the two surveys, and for different types of questions in the Interview survey, may be different. Because of these differences, we cannot conclude whether a recall interview or a diary is inherently a better mode for obtaining the most accurate expenditure data across a wide range of items. Both have real drawbacks, and a new design will need to draw from the best (or least problematic) aspects of both methods.

SOURCES OF RESPONSE ERROR IN THE INTERVIEW SURVEY

The CE Interview survey is long and exhausting. A household respondent is expected to complete this interview five times, three months apart. The interviews average 60 minutes but may be shorter or much longer. (Panel members who reported their own expenditures in mock interviews with Census field representatives described interviews that lasted significantly longer.) During the interview, respondents are asked to report as many as 1,000 specific expenditures during the preceding three months. These questions cover the gamut of items for which a household might expend dollars, including health insurance, women’s blouses, children’s toys, men’s socks, toasters, the repair of an air conditioning unit, vehicle cleaning, mortgage interest, electricity, prescription medications, alcohol, gasoline, greeting cards, and parking and tolls, to mention just a few. To put the enormity of the task in perspective, an information booklet is handed to the respondent at the beginning of the interview. This booklet includes

36 pages with 9 to 70 items per page of possible consumer expenditures the respondent is asked to report.

Inaccurate reporting to this gauntlet of questions will occur. The rest of this section highlights some of the potential reasons for these errors.

Motivation in Interview Survey

Respondents in the CE Interview have little apparent motivation to engage in a complex, protracted interview. Once a household member agrees to participate in the CE Interview survey, he or she discovers that the task is cognitively difficult and time-consuming. Some respondents see the reporting of detailed expenditures as an invasion of privacy. Others may fear sharing certain information with a government agency. Some expenditures, such as gambling losses or excessive purchases of alcohol, may be embarrassing to report. Beyond those concerns, the majority of respondents just want the interview to be over as quickly as possible (Mockovak, Edgar, and To, 2010).

The field representatives understand these concerns. When asked about factors that contributed to underreporting on the CE, they said the greatest factor was *respondent mental fatigue because the interview is too long*. Sixty-two percent (62 percent) rated this factor as a 6 or 7 on a 7-point scale of importance (Mockovak, Edgar, and To, 2010). Field representatives want to do their job: complete the current interview and return to repeat the process four more times. They feel a need to keep the interview short so that it does not end up as a refusal, either immediately or in subsequent waves. This produces a tradeoff between completing the interview and pushing too hard for accurate answers.

There is little doubt that *both* the respondent and the field representative benefit from keeping the interview short. Encouraging respondents to give more complete answers, for instance by encouraging them to find and consult records or consult with other household members, is likely to slow down the interview process. Placing an emphasis on getting exact amounts is also likely to lengthen the completion process and frustrate respondents.

Some respondents may be initially motivated to report accurately but soon find that they cannot. Accurate recall of the hundreds of items on the CE is very difficult. Even motivated respondents may find they are not able to do this. (See panel members' reactions to completing the CE in Chapter 4, in the section "Panelists' Insight as Survey Respondents.") Under these conditions the motivations of both the respondent and field representatives affect the accurate reporting of expenditures.

Conclusion 5-3: Motivational factors of both the respondent and field representative appear to negatively influence the quality of the CE

Interview data. This leads the panel to the judgment that a changed incentive and support structure for both respondents and field representatives will be needed for a future CE redesign to motivate high-quality reporting and reduce fatigue.

Interview Questionnaire Structure

The current CE Interview questionnaire is structured around categories of expense items. The field representative asks first about a fairly broad category of items and then drills down until the question is directed toward a specific detailed item. For example, we will ask first about any clothing purchases: “Did you purchase any pants, jeans, or shorts?” At this point, the questionnaire asks a series of ancillary questions about the purchased item.

- *Describe the item.*
- *Was this purchased for someone inside or outside of your household?*
- *For whom was this purchased? Enter name of person for whom it was purchased. Enter age/sex categories that apply to the purchase.*
- *How many did you purchase? Enter number of identical items purchased.*
- *When did you purchase it?*
- *How much did it cost?*
- *Did this include sales tax?*
- *[if the respondent cannot separate the item from other items] What other clothing is combined with the item? Enter all that apply from a list of 18 clothing categories.*

The questionnaire then returns up one level of aggregation to identify other expenditures within that subcategory (“Did you purchase any other pants, jeans, or shorts?”). This questionnaire structure creates the cognitive challenges described below.

The Structure of the CE Interview Encourages Satisficing and Similar Response Errors

The CE Interview asks a series of *global questions* that require respondents to think about unnamed specific items (e.g., pants, socks, belt) they have purchased based upon a general stimulus (e.g., clothing). If they answer “yes” to the global question, then they will be asked *specific questions* about that purchase or purchases. This sequence is repeated dozens of times during each interview and may affect respondent behavior. It seems likely

that respondents learn quickly in the first interview, and are reminded in each successive one, that the interview will last longer if they answer “yes” to these screening questions. For example, a respondent would be asked if anyone in the household took any trips during the three-month period. A “yes” answer leads to many questions about specific expenditures made on that trip. After completing that series of specific questions, the respondent is then asked if household members took any other trips. The respondent quickly understands that reporting a second trip would add a number of additional questions and minutes to the interview. This phenomenon is known as “motivated underreporting” and is discussed by Kreuter et al. (2011). A survey of CE field representatives (Mockovak, Edgar, and To, 2010) quantified this problem in the CE. Field representatives were asked how often this phenomenon happens in a CE interview. Fifty percent of field representatives said that it happened *frequently* or *very frequently*.

Conclusion 5-4: The current structure of the Interview questionnaire cycles down through global screening questions, and asks multiple additional questions when the respondent answers “yes” to a screening question. As this cycle repeats itself, a respondent “learns” and may be tempted not to report an expenditure in order to avoid further questions.

CE Methods Are Not Well Aligned with Modern Consumption Behavior

The CE Interview questionnaire is cognitively designed to collect spending information in an earlier era when purchases and expenditures were made in quite different ways. The cognitively outdated design of the questionnaire makes it difficult for consumers to respond easily and accurately to the questions. This exacerbates both recall error problems and overall response.

Major changes have occurred in retail markets in the last decade, including a major consolidation of the retail pharmacy and grocery industry. Simultaneously there has been an explosion of loyalty card programs in these same (and other) industries. These days, it is common for households to purchase a variety of types of items in a single large store, such as Costco or Walmart, rather than going separately to a grocery store, butcher shop, clothing store, and hardware store. A single purchase of a group of items and the payment of a “total amount” may make it more difficult for a consumer to later recall details about an individual item than if that item were purchased in a separate transaction. Separate transactions provide a focus on the individual items and may help reinforce the memory of both purchasing the item and the amount paid.

Consumers purchase items through many different methods, including credit card, debit card, check, cash, gift card, payroll deduction, and preauthorized automated payment. Items may be purchased in person, by telephone, by mail, or online. Respondents may remember how much they paid on their credit card bill but be unable to recall the specific items that were purchased. Respondents may not think at all about automatic payments. The combined effects of these increasingly varied ways of making purchases and a rigid interview questionnaire that generally flows by product groupings rather than by shopping trip or payment method make the task of recalling and reporting those expenditures more difficult.

This question structure seems likely to encourage the use of “estimation” rather than the reporting of a specific recall, and ultimately may lead to less accurate reporting of particular expenditures (Beatty, 2010; Peytchev, 2010). Some questions on the Interview survey (such as *food at home*) specifically ask the respondent for estimates rather than specific recall. Other questions ask for a specific recall, yet it is unclear in these questions how much estimation is also taking place.

Conclusion 5-5: The current design of the CE Interview questionnaire makes the cognitive task of recalling expenditures difficult and encourages estimation.

Some Questions Are Just Difficult to Answer

In the fifth interview, respondents are asked a series of questions about household financial assets:

- *On the last day of last month, what was the total balance or market value (including interest earned) of checking accounts, brokerage accounts, and other similar accounts?*
- *On the last day of last month, what was the total balance or market value (including interest earned) of U.S. savings bonds?*
- *How does the amount your household had on the last day of last month compare with the amount your household had on the last day of last month one year ago?*

These questions and others like them that ask for precise accountings by month are very difficult for respondents to answer. Banks are likely to provide monthly statements for checking accounts, but holders of savings and other asset accounts are generally provided with quarterly statements. A respondent is unlikely to know the market value of those accounts on the last day of the last month unless that day corresponds with a quar-

terly statement. Think of the frustration of respondents who do not have the information to answer that question accurately and are then asked to compare their estimate with the market value of those same assets one year earlier. Regarding U.S. savings bonds, individuals may monitor the maturity date of those bonds but are very unlikely to observe the growth in value on a monthly basis, nor even know how to do so.

Respondents are asked for the purchase date of most expenditures they recall. For some, they may be confused about the date on which a particular “expense” occurred. This can be particularly problematic with online or mail purchases. Was it “purchased” when the order was placed, when the item arrived, or when the bill was paid? In the urgency to complete the interview quickly, these CE guidelines may not be explained, understood, or remembered.

Some consumer transactions occur quickly and routinely without the purchaser remembering the cost, even momentarily. Specific purchases and prices that did not mentally register with the respondent cannot be reported later (Bradburn, 2010). Imprinting the “event” in memory or encoding, as psychologists describe it, is less likely to happen with minor, routine purchases. The use of credit and debit cards to pay for groups of varied purchases in large stores seems less likely to result in encoding for specific purchases and prices. In addition, credit or debit cards are increasingly likely to be used for even small routine purchases (lunch, a newspaper, or garage parking), contributing further to the lack of encoding. Automatic deductions of payments from a bank account may also contribute to a lack of encoding. Thus, multiple aspects of contemporary society appear to increase the difficulty of respondents’ ability to report expenditures accurately or at all.

Even if respondents remember how much they paid for a shirt, they may have difficulty knowing whether the amount included sales tax. Additionally, many online purchases are made without the inclusion of state sales tax. The respondent may answer that the amount of an online purchase did not include sales tax, but the CE process will then add tax to that purchase when no tax was actually paid.

Conclusion 5-6: Some questions on the current CE Interview questionnaire are very difficult to answer accurately, even with records.

Interview Survey Recall Period

The CE Interview questionnaire asks respondents to recall most expenditures over the previous three months. The issue of recall, and its effect on reporting accuracy in the CE, was a major topic in the BLS-sponsored

CE Methods Workshop in December 2010. Cantor (2010, p. 4) provided a basic summary, saying “longer recall periods lead to more measurement error. For the CEQ [CE Interview], there are two important characteristics related to error. One is whether the expense is reported at all. The second is the detail associated with the event.” A BLS paper at that same workshop expressed concern:

The length of this three-month recall period, combined with the wide range of question types asked, is generally thought to represent a substantial cognitive burden for respondents. Furthermore, there are different approaches to asking about the three-month recall period, which may compound the cognitive burden for respondents. For example, some CEQ [CE Interview] questions ask about cumulative expenses over the entire three-month recall period, other questions ask respondents about total monthly expenditures for the first, second, and third month of the recall period, and still others ask respondents for average weekly expenses over the recall period. (Bureau of Labor Statistics, 2010c, p. 1)

A three-month recall period can be appropriate for major expenditures such as a major appliance or an automobile, or for expenditures that occur on a regular basis such as rent and utility bills. These are the types of purchases for which the CE Interview survey was initially designed (Bureau of Labor Statistics, 2008). However, the survey currently attempts to collect data on expenditures less likely to be remembered. It also asks detailed information about those expenditures that are difficult to recall. A common response error is one of omission—to simply not remember or not report the purchase and/or price of a particular item. These errors appear to be a major factor in the underreporting of expenditures on the CE.

On the other hand, a longer reporting period (be it for recall or concurrent reporting) has some advantage for estimation. Frequent expenditures may be captured well by a short reporting period, while other expenditures are less frequent and will not be reported by all households during a short recall period. A sufficient sample size collected throughout the year will avoid bias in the estimates, but there is likely to be more variability in the estimates with a shorter reporting period if all other factors are equal. If infrequent expenditures are also major expenditures (i.e., easily remembered), then a longer recall period for those items may be reasonable.

Conclusion 5-7: Three months is long for accurate recall of many items on the CE Interview survey. This situation is exacerbated by the ancillary details that are collected about each recalled expense. Errors of omission are likely to occur, and are a contributing factor to the underreporting of expenditures on this survey. Short recall periods, however,

may produce more variability in the estimates and provide difficulties for economic research.

Use of Records in the Interview Survey

Some respondents do not keep records for expenditures, nor do they keep receipts that specify item amounts. If they have records, they may only keep grand totals without breakouts by item. When records are available, they likely are not organized in a way that allows them to be used effectively (or efficiently) during the interview. Field representatives have reported that a respondent occasionally goes in search for a receipt or record, to return sometime later without it.

Respondents who keep electronic records may not have them up to date in their computer when the interview is conducted and therefore cannot use them to provide accurate answers. A respondent's electronic records are most likely organized differently from how they are requested in the interview. For example, a respondent may have a record of automobile fuel costs, but may not keep it separately for different cars; purchases at Safeway may not be broken down by food and other household items, all being considered "groceries."

Paper receipts may be problematic as well. Groups of purchases are likely to be combined into one receipt, and that receipt may be difficult to decipher. Coupon discounts, discounts for the use of a particular credit card, or special "in-store" sales may be shown separately on receipts, so even a conscientious respondent may have difficulty understanding what was actually paid for a specific item. The receipt probably does not include rebates that consumers receive at a later point in time. Receipts vary enormously across stores, with abbreviations, special coding, and character limitations affecting how a particular purchase is described. The respondent may not be able to interpret that information, especially as the memory of the specific purchase fades over days and weeks.

BLS conducted a study on use of records for the CE, using data from CE interviews conducted between April 2006 and March 2008. Following each interview, the field representative recorded whether the respondent used records and which types of records were used. Edgar (2010) indicated that the study involved 44,300 interviews with 21,011 unique households. She reported that in 39 percent of interviews, the respondent *never* or *almost never* used records, while 31 percent *always* or *almost always* used records. Younger respondents were less likely to use records than older respondents. Interviews were longer when records were used and shorter when they were not used. There was a higher level of total expenditures reported when records were used. Those *always* or *almost always* using re-

cords as a group reported a higher level of expenditures. In summary, Edgar (2010, p. 26) found that the use of records is “related to longer interviews, more reports, and higher reports.”

These patterns appear again in a later (2010) evaluation survey. Field representatives working on the CE surveys were asked how often respondents use records and receipts to help report their expenditures. Thirty-two percent reported that respondents *rarely* used these records and another 50 percent reported that respondents only *sometimes* used records. Only 18 percent of field representatives said that respondents *often* used records. In that same study, 68 percent of field representatives said that respondents *never* or *rarely* consulted online or electronic records.⁷

Geisen, Richards, and Strohm (2011) reported on a study conducted for BLS on the availability of records and response accuracy without those records. They conducted two interviews with the same households ($n = 115$) within a week of each other. Incentives were given for both interviews. The first interview asked respondents to complete nine sections of the CE Interview survey. Afterward, respondents were asked to gather all of the records they had relevant to that interview, and they were re-interviewed within the week with their gathered records. The second interview focused on matching the original response to available records. They found that respondents had records for only 36 percent of the expenditure items and 41 percent of the income items. Records were most likely to be available for property taxes (59 percent), mortgage payments (59 percent), and subscriptions (53 percent). The most recent pay stub was available only 40 percent of the time. Regarding response accuracy without records, the authors found that the expenditure amount reported in the first interview “matched”⁸ the amount on the record only 53 percent of the time. The difference varied greatly by expenditure categories, and ranged from an 11 percent underestimate to an 83 percent overestimate.

The increasing use of telephone interviewing in the CE may reduce the respondent’s use of records in responding. In preliminary results from a 2011 survey of field representatives working on the CE survey (Bureau of Labor Statistics, 2011d), 45 percent of field representatives reported that respondents were *much less* or *somewhat less* likely to use records when being interviewed over the phone. Only 4 percent of field representatives reported that respondents were *much more likely* or *somewhat more likely*

⁷Internal Bureau of Labor Statistics memo dated November 18, 2010.

⁸Geisen, Richards, and Strohm (2011) classify a matched response as one in which the initial response was between 90–110 percent of the record amount (for amounts \leq \$200); 95–105 percent of the record amount if the amount was $>$ \$200; and 95–105 percent of the record amount for rent, mortgage loan, and income regardless of amount.

to use records in a phone interview. However, 39 percent of field representatives reported that respondents were *neither less likely nor more likely* to use records on a phone interview.

Conclusion 5-8: The use of records is extremely important to reporting expenditures and income accurately. The use of records on the current CE is far less than optimal and varies across the population. A redesigned CE would need to include features that maximize the use of records where at all feasible and that work to maximize accuracy of recall when records are unavailable.

Proxy Reporting in the Interview Survey

According to the CE procedures, multiple household members are not interviewed even though most households have more than one person who makes purchases. The field representative attempts to interview the “most knowledgeable” household member. Schaeffer (2010) reported that proxy information reported by one person for another in many types of surveys has been found to be inaccurate. In the CE, the household respondent may be unaware of purchases made by other household members (Bureau of Labor Statistics, 2010b).

The effect of proxy reporting is serious, especially for “personal” purchases such as clothing or for purchases that household members want to keep private. A survey of CE field representatives (Mockovak, Edgar, and To, 2010) quantified this problem. Field representatives reported the issue of the household respondent not knowing about purchases made by others in the consumer unit (household) as one of the two most important reasons for underreporting of expenditures. More than half (53%) mentioned it as 6 or 7 on a 7-point scale of importance. In households with multiple purchasers, field representatives reported 18 percent of the time that a second person *never or almost never* participates in the CE survey. An additional 48 percent mentioned that it happens *less than half of the time*.

Multiple considerations may underlie the inability of one person to report for others. Household members may have separate sources of income. In addition, they may have separate bank accounts and credit cards. While each person may contribute to household expenses, the exact amounts of the other person’s contribution may be unknown; for example, one person buys food while the other pays housing costs. The great variety of arrangements for handling individual and joint income complicates the reporting of consumer expenditures. In addition, field representatives have reported it appears that some other household members may intentionally withhold information from the household respondent.

Conclusion 5-9: The use of proxy reporting on the CE Interview is problematic, and is a potential cause of underreporting of expenditures.

Telephone Data Collection in the Interview Survey

More than one-third (about 38 percent) of the CE interviews are completed by telephone rather than face-to-face (Safir and Goldenberg, 2008). For in-person interviews, field representatives hand an *Information Booklet* to respondents to assist them in recalling items that might be forgotten. In telephone interviews, this recall aid is not available. Telephone interviews are shorter and have fewer positive answers to screener questions. They also result in less detail in responding, higher item nonresponse, and more reporting of rounded values. According to the field representative survey, receipts and other records are also less likely to be used by telephone respondents.

Conclusion 5-10: Telephone interviews appear to obtain a lower quality of responses than the face-to-face interviews on the CE, but a substantial part of the CE data is collected over the telephone.

SOURCES OF RESPONSE ERROR IN THE DIARY SURVEY

The Diary survey collects data on expenditures a household makes during a brief period of time (two weeks). It was designed to collect a level of detail unlikely to be recalled accurately during the Interview survey. Cognitively, the Diary survey is very different from the Interview survey, and has its own error profile.

Motivation to Complete the Diary

The current CE Diary is designed to reduce recall problems by emphasizing the recording of expenditures the day they are made. However, respondents in the CE Diary have little apparent motivation to engage in a complex, protracted diary-keeping exercise. Once a household member agrees to participate in the CE Diary survey, he or she discovers that the task can be difficult and time-consuming. Some respondents do not record expenditures each day. Expense items may be reported early in the week, with less enthusiasm later in the week, and even less enthusiasm during the second week. This lapse may be caused by general fatigue with the process and/or the fact that the respondent found the diary form difficult to use. Statistics Canada found the same concern with the diary portion of their consumer survey, reporting that 20 percent of their two-week diary days were “nonrespondent” days (Dubreuil et al., 2010). This lack of motivation

of a respondent to stop in the middle of busy daily activities and record an expenditure on the diary form is probably the major cause of underreporting of expenses in the current Diary survey.

Field representatives place mid-week calls to diary-keepers to mitigate this problem—to ask whether the diary-keeper is having any difficulties and to encourage continued reporting through the week. In addition, the diary pick-up at the end of each week provides the field representative an opportunity to explore whether the respondent may have forgotten to record certain expenditures such as those from checks, cash, credit card payments, automatic online payments, or deductions from pay stubs. These mitigation strategies are not uniformly implemented. There are shortcuts allowed in the fielding of the Diary survey; for example, field representatives are permitted to place both one-week diaries at the same time. Thus, no “reinforcing” visits take place in these cases to examine diary entries and to encourage increased compliance during the second week. The panel did not have data on the extent of this practice or on how often the mid-week telephone calls were placed with households.

Conclusion 5-11: A major concern with the Diary survey is that respondents appear to suffer diary fatigue and lack motivation to report expenditures throughout the two-week data collection period and especially to go through the process of recording all items in a large shopping trip.

Diary Structure

The current diary form suffers from a number of cognitive problems that make the diary-keeping process more difficult than it needs to be and thus can contribute to underreporting on those forms.

Learning to Complete the Diary May Be Confusing

An initial challenge facing respondents willing to complete the diary is sorting through all the instructions to understand how and where to report expenditures. The diary has 44 numbered pages plus front and back covers, both of which have attached foldout flaps. Fifteen separate pages (including covers and foldout flaps) provide instructions for the diary-keeper on what and how to report, with instructional material scattered among these 15 pages. On the front cover, the field representative identifies the days the diary is supposed to be kept and the first names of the people in the household. The foldout cover flap lists numerous examples for each category of expenditures. Page 1 tells “Why the CE Diary Is Important.” Page 2 describes “How to Fill Out Your Diary” and the four parts. Pages

4 through 7 provide examples of how to fill out each of the four sections. Finally, the back cover provides a pocket and instructions for storing receipts and other expenditure records. It also includes a Daily Reminder list of 18 items with an instruction to ask other members of the household for their expenditures each day. Another foldout flap provides answers to 15 frequently asked questions.

The diary booklet is not organized so as to reveal a natural linear process for becoming acquainted with the recording process. The field representative must flip through the booklet to train the respondent on its use, referring to the 15 individual instructional pages (including foldout flaps) to explain how the diary is designed to be completed. Although each field representative is likely to handle the diary placement somewhat differently, the following may be a typical set of instructions to a potential diary-keeper. Figures 5-2 through 5-5 show four relevant pages from the diary:

- Turn to page 8 (instead of starting at page 1). This page is labeled “Day 1” with a separate heading labeled “1. *Food and Drinks Away from Home.*” Circle the “day of week” that appears in a third heading.
- Moving to the substance of the page, the field representative might explain the six columns of information that need to be answered for each food entry: type of meal; description of meal; where was meal purchased; total cost; which of three types of alcoholic beverages might have been purchased; cost of that alcohol.
- Turn to page 3, to a section on “*How to Fill Out Your Diary.*” The field representative explains that the food and drinks section previously reviewed was only one of four such tables that have to be filled out each day.
- Turn to page 4. This page displays detailed examples of how the food and drinks might be reported.
- Move back and forth between blank diary pages, examples of filled-out pages, and other sections. These sections include general instructions (page 2), keeping receipts in the back pocket of the diary, a “What Not to Record” section, and, on the back flap of the cover, answers to 15 different questions that the respondent might have.

Conclusion 5-12: A lot of information is conveyed to the diary respondent in a short amount of time. The organization of the diary booklet may result in considerable frustration among some individuals, who feel they cannot master the instructions. They choose instead to collect receipts and leave them for the field representative to enter during the follow-up visit.

Day 1					SUN	MON	TUE	WED	THU	FRI	SAT		
1. Food and Drinks Away from Home													
Examples:		breakfast buffet carry-out lunch dinner & cocktails at restaurant	pizza delivery Chinese takeout child's school lunch	beer at happy hour pretzels at ballgame wine at tavern	croissant from café ice cream from truck wedding reception caterer	soda from vending machine hot dog from convenience store popcorn and soda at movies	Please unfold the LEFT FLAP to see Additional Examples						
Mark (X) one that best describes the type of meal	Description <i>(see examples above and on the flap)</i>				Mark (X) one that best describes where you made this purchase				Total Cost with tax & tip	If alcoholic beverages included, mark (X) all that apply			Enter the total cost of the alcohol
					Fast-Food Take-out Delivery Concession	Full Service Places	Vending Machines or Mobile Vendors	Employer or School Cafeteria		wine	beer	other	
breakfast	lunch	dinner	snack/other	1	2	3	4						
101													
102													
103													
104													
105													
106													
107													
108													
109													
110													
111													
112													
113													
114													
115													
116													
117													
118													
119													
120													
121													
122													
If there are not enough lines in this part, please continue recording your expenses on pages 36-37.													

8

FORM CE-801 (7-1-2005)

FIGURE 5-2 Diary booklet, page 8.
SOURCE: Bureau of Labor Statistics.

How to Fill Out Your Diary

The diary is divided into 7 days and each day is divided into 4 parts. Enter each item in the appropriate part for each day.

These are the 4 parts within each day of the diary:

1. Food and Drinks Away from Home

- Mark one of the four choices that best describes the type of meal and describe briefly.
- Mark one of the four choices that best describes where you made the purchase.
- Enter the total cost **with tax and tip**.
- If alcohol was part of the purchase, check whether it was wine, beer, and/or other alcohol and enter the total cost of the alcohol.

2. Food and Drinks for Home Consumption

- Describe the item.
- Mark whether the item was fresh, frozen, bottled/canned, or other.
- Enter the cost **without tax** and deduct any discounts or coupons.
- Mark the last column if the item was purchased for someone not on your list (e.g. gifts).

3. Clothing, Shoes, Jewelry, and Accessories

- Describe the item and enter the cost **without tax**.
- Mark the appropriate sex and age range of the person for whom the item was bought.
- Mark the last column if the item was purchased for someone not on your list (e.g. gifts).

4. All Other Products, Services, and Expenses

- Describe the item and enter the total cost **without tax**.
- Mark the last column if the item was purchased for someone not on your list (e.g. gifts).

There is an "Additional Pages" section on pages 36–44 in case you run out of lines on any particular day.

Look on the next 4 pages for examples and tips on how to record your purchases.

***Please Note: If you are unsure about whether to include an item or where to record an item, write it down wherever it seems best or make a note and ask your field representative.**

FIGURE 5-3 Diary booklet, page 3.
SOURCE: Bureau of Labor Statistics.

EXAMPLE					SUN	MON	TUE	WED	THU	FRI	SAT					
1. Food and Drinks Away from Home																
Examples: breakfast buffet carry-out lunch dinner & cocktails at restaurant pizza delivery Chinese takeout child's school lunch beer at happy hour pretzels at ballgame wine at tavern croissant from café ice cream from truck wedding reception caterer soda from vending machine hot dog from convenience store popcorn and soda at movies																
← Please unfold the LEFT FLAP to see Additional Examples																
Mark (X) one that best describes the type of meal	Description (See examples above and on the flap)				Mark (X) one that best describes where you made this purchase					Total Cost with tax & tip	If alcoholic beverages included, mark (X) all that apply			Enter the total cost of the alcohol		
					Fast Food Take-out Delivery Concession	Full Service Places	Vending Machines or Mobile Vendors	Employer or School Cafeteria	Include tax & tip for part 1 only.		wine	beer	other			
breakfast 1 2 3 4	lunch 1 2 3 4	dinner 1 2 3 4	snack/other 1 2 3 4	Level of detail needed: briefly describe the meal.												
101	X								X	2	79					
102		X				X				5	57					
103			X			X				1	35					
104		X							X	5	15					
105			X						X		70					
106		X							X	45	00					
107			X						X		65					
108			X				X			62	23	X		12	00	
109			X				X			15	00		X	X	15	00
110			X				X			350	00	X	X	X	95	00
111																
112																
113																
114																
115																
116																
117																
118																
119																
120																
121																
122																

If there are not enough lines in this part, please continue recording your expenses on pages 36-37.

4 FR USE: None TR VC FORM CE-801 (7-1-2005)

FIGURE 5-4 Diary booklet, page 4.
SOURCE: Bureau of Labor Statistics.

General Instructions

- Fill out this diary for an entire week, writing down EVERYTHING you and the people on your list spend money on each day – the products you buy, the services you use, the household expenses you have during the week – no matter how large or small they are.
- We recommend that you record your expenses each day. Think about where you went and what you've done.
- Talk to the people on your list every day to find out how they spent their money.

- Include payments by:

Cash	Credit/Debit Card	Automatic Withdrawal/Payroll Deduction
Check	Money Order	Store Charge Card
Food Stamps	WIC Voucher	Grocery Certificate

- Keep receipts and other records so that you will remember to record what you bought or paid for. Use the pocket at the back of the diary to store them.

Some record types include:

Receipts	Bank Statements	Catalog/Internet Order Invoices
Utility Bills	Telephone Bills	Credit Card Statements
Pay Stubs		

Include items that you bought for people who are not on your list, such as gifts.



Do NOT record:

- ◆ Expenses of people on your list while they were away from home overnight.
- ◆ Business or farm operating expenses
- ◆ Sales tax for:
 - Part 2. Food and Drinks for Home Consumption
 - Part 3. Clothing, Shoes, Jewelry, and Accessories
 - Part 4. All Other Products, Services, and Expenses

FIGURE 5-5 Diary booklet, page 2.
SOURCE: Bureau of Labor Statistics.

Recording Expenditures May Be Problematic

The diary instructions focus on recording expenditures each day during the diary week, separately for different categories. A total of 28 pages of the booklet are laid out by “day” and consist of labeled tables for recording household expenditures made in each of four categories:

1. Food and Drinks Away from Home
2. Food and Drinks for Home Consumption
3. Clothing, Shoes, Jewelry, and Accessories
4. All Other Products, Services, and Expenses

Nine additional pages are included for reporting any information that will not fit on the individual day pages.

The diary-keeper is to record each expenditure on the correct page by day and expenditure category. For each item recorded, the form asks for a description of the item, the cost, plus additional information that differs for each of the four expenditure categories. For example, the tables for *Clothing, Shoes, Jewelry, and Accessories* ask for additional information on the individual for whom the item was purchased: gender, age, and whether the person was a member of the household. The tables for *Food and Drinks Away from Home* also ask where the item was purchased, whether the item included alcoholic beverages, and a breakout of the cost of any such alcohol.

Research has shown that respondents are influenced by much more than words on how to complete questionnaires; a mostly linear path with guidance from numbers, graphics, and symbols also helps to instruct respondents on how a questionnaire (or diary) is designed to be completed (Dillman, Smyth, and Christian, 2009). The current in-person delivery and retrieval process is designed to compensate for some of these problems. The field representative may look at the receipts collected by the household respondent and ask other questions to find out whether all expenditures have been recorded. Realization at the initial visit that the interviewer will call mid-week and return to pick up the diary at the end of the week would seem to encourage respondents to think about their daily expenditures and be able to recall and report them at the end of the diary week contact. However, this mitigation process is not always followed.

Other general problems can occur with the diary. Some shopping trips require complex reporting of many and varied items on the diary forms. For example, a major grocery-shopping trip may take considerable time and effort to record. Each item purchased must be itemized separately with a description and cost. (The respondent has to figure out whether to record before or after the food is put away.) Receipts are often limited to abbrevia-

tions and codes that may not be understandable to the person who made the purchase and/or is completing the diary.

Part of the diary placement visit is for the field representative to “size up” the respondent as to whether he or she understands how to complete the diary and seems committed to do so. If the respondent does not appear to understand the instructions and the use of the daily recording forms, some field representatives will revert to an alternative approach and ask such a respondent to merely keep all of the household receipts for the week’s expenditures in a pocket of the inside back cover. On the next visit a week (or two weeks) later, the field representative and the respondent will go through the receipts and fill in the diary forms together. The panel learned that this approach is likely used for a significant number of households in which the respondent finds keeping the diary too difficult to do.

Conclusion 5-13: It is likely that the current organization of recording expense items by “day of the week” makes it more difficult for some respondents to review their diary entries and assess whether an expenditure has been missed.

Reporting Period for the Diary Survey

The Diary survey has a one-week reporting period, followed immediately by a second wave also consisting of a one-week reporting period. The Diary survey was conceived as a vehicle for collecting smaller and frequently purchased items that were unlikely to be reported accurately over a three-month recall period. However, in practice, the Diary collects a wide variety of expenditure items. Since many types of expenditures are made infrequently, and others are not purchased in the same amount each week, Diary expenditure estimates for these variables are likely to be more variable than those from the Interview survey with its three-month reporting period. For example, Bee, Meyer, and Sullivan (2012) found that for 2010 the weighted average coefficient of variation of spending reports on 35 categories of expenditures common to both the Interview and Diary was nearly 60 percent higher for a typical Diary response than for a typical Interview response.⁹ One reason is that, in 2010, close to 10 percent of weekly diaries that were considered as valid observations reported no in-scope spending at all. (About 75 percent of these reports were out-of-scope because the family was on a trip for the week.) Consequently, a larger number of weekly diaries is required to equal the precision of the quarterly interviews.

⁹Their comparisons adjusted for the different sample sizes of the Diary and Interview surveys. The coefficient of variation is the standard error of a mean or other statistic expressed as a percentage of the statistic.

A related conceptual issue is that the short reference period in the current Diary survey may be too short to accurately measure an individual's normal spending pattern. While these errors may average out in the calculation of means, several important uses of the CE require the measurement of the distribution of spending.

Proxy Reporting in the Diary Survey

Respondents are asked to consult with other members of the household during the week and to report expenditures for all members. The field representative lists the names of the members on the inside cover of the diary. These instructions are aimed at encouraging communication between the person who agrees to complete the diary and others in order to facilitate accurate reporting. As stated earlier, the diary mode does provide more opportunity to confer with other household members over the week than there is within a rushed recall interview. However, there are still issues with proxy reporting in today's households.

The changing structure of U.S. households, in which the adults in that household are more likely to have separate incomes and expenditure patterns, means that unless deliberate communication occurs expenditures may be underreported. In addition, household members are not always open with each other about what they have purchased or how much it cost. If a member of the household does not want the person most responsible for completing the diary to know of the expenditure or its cost (e.g., a teenager downloading a new video game, the cost of an anniversary gift, or payment of a parking ticket), the diary will probably miss the expense.

Conclusion 5-14: Although the diary protocol encourages respondents to obtain information and record expenditures by other household members during the two weeks, it is unclear how much of this happens.

NONRESPONSE

Comparison of Response Rates

In calculating response rates on the CE Interview survey, BLS uses outcome information from each household for each wave (waves two through five) as independent observations in the survey. For the Diary survey, BLS counts each week of the two weeks of diary reporting by a household as an independent observation. The "CE program defines the response rate as the percent of eligible households that actually are interviewed for each survey" (Johnson-Herring and Krieger, 2008, p. 21). These calculations exclude cases where the household is ineligible.

BLS sometimes refers to the rates described by Johnson-Herring and Krieger as the “collection rates” since they are computed by the Census Bureau immediately following data collection. In post-collection processing, BLS removes some data records because they are found to have limited data entries and expenditure dollar amounts. BLS then recalculates the response rates, and sometimes refers to these adjusted rates as “estimation rates.” Johnson-Herring and Krieger do not discuss these adjustments in their description of methodology. In comparing the “collection rates” with the “estimation rates” one sees that the adjustments affect the Diary rates more than the Interview rates. BLS generally provides the “estimation rates” as the response rates to their microdata users. After some consultation with BLS, the panel has concluded that the adjusted “estimation rates” more closely describe the usable response, and has decided to use those as the response rates for the purpose of this report.

Thus, as reported in Chapter 3, the CE Interview survey had a response rate (estimation rate) in 2010 of 73 percent, slightly ahead of the Diary survey, which had a response rate of 72 percent.

Both surveys have experienced declines in response rates over time, a problem that has plagued most government household surveys. The response rates (estimation rates) for the Diary survey have been slightly lower than those for the Interview survey (see Figure 5-6). The CE is a burdensome survey, and the overall response rate is lower than several well-known but less burdensome surveys such as the Current Population Survey (92%) and the CPS Annual Demographic Survey (80% to 82%).

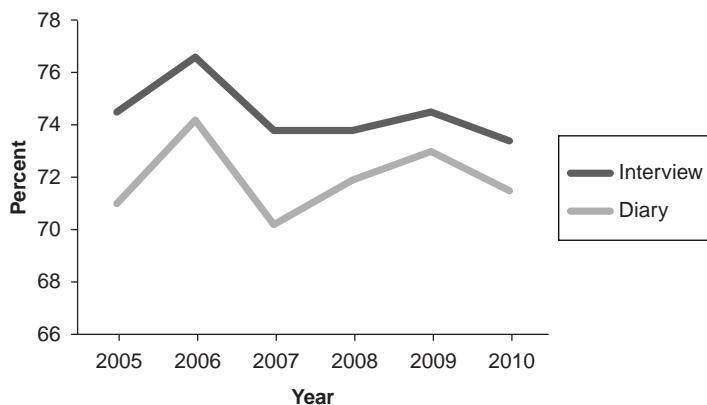


FIGURE 5-6 Response rates (estimation rates) for Consumer Expenditure Interview and Diary surveys.

SOURCE: Bureau of Labor Statistics, data table provided to panel.

However, the CE's response rate is comparable to consumption surveys in other countries, which have experienced similar declining response rates during the period between 1990 and 2010. Figure 5-7 depicts the response rate to the CE (United States) compared to response rates for comparable consumption surveys in Australia, Canada, and the United Kingdom (Barrett, Levell, and Milligan, 2011). The CE response rate is somewhat higher than the others.

Panel Attrition

Given that the CE Interview survey uses a rotating panel design with five waves of data collection per panel, an additional response concern relates to attrition over the life of a panel (Lepkowski and Couper, 2002). King et al. (2009) studied the pattern of participation in the CE. In this study, they looked at a single cohort first interviewed in April–June 2005. Among the households that completed the first wave interview, 78.6 percent were classified as *complete responders* (data for all five waves were captured), 14.1 percent were classified as *attriters* (completed one or more

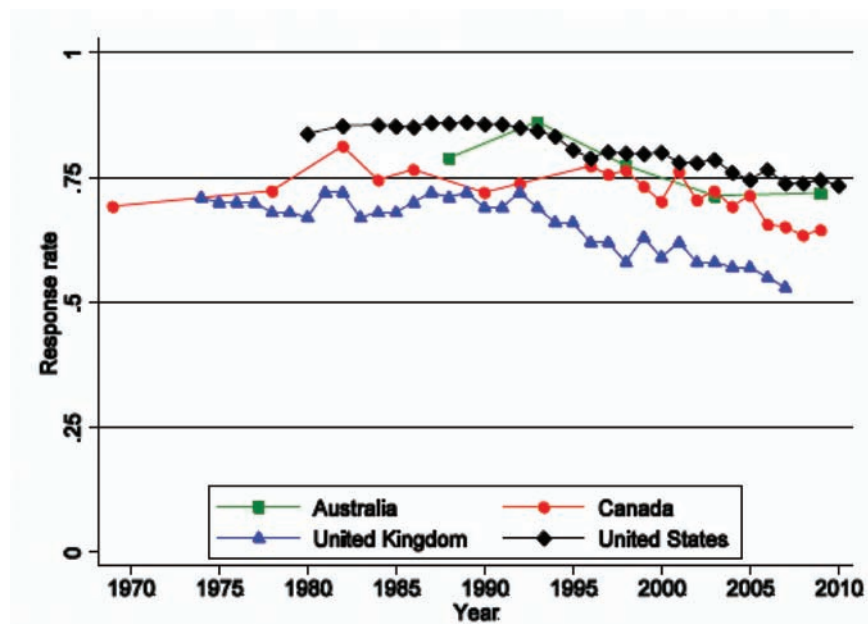


FIGURE 5-7 Response rates for consumption surveys across four western countries. SOURCE: Barrett, Levell, and Milligan (2012).

interviews before dropping out), and 7.3 percent were classified as *intermittent responders* (completed at least one but not all waves of data collection). It is not clear whether the response rates reported above are for all cohorts in a particular quarter (i.e., averaging across different panels), but the drop-off after the first wave of data collection raises both concerns and opportunities.

A key concern is that the decision to attrite may be influenced by the experience of the first wave, and that experience may be different depending on the number and types of expenditures reported in that wave. If this is the case, the attrition can potentially affect the estimation of expenditures during later waves. In other words, to what extent is later wave participation affected by reported expenditures (or expenditure patterns) in the first interview, holding constant the usual demographic variables that are used in weighting adjustments? The panel is not aware of research exploring this potential source of bias.

A key opportunity for the future: BLS can use households that provide partial data (i.e., attritors or intermittent responders), along with their level and pattern of expenditures, in the adjustment for missing waves. It is the panel's understanding that the nonresponse adjustments employed by BLS use post-stratification adjustment to population controls in a cross-sectional adjustment and do not make use of the household's expenditure data from other interviews. While this may be hard to do given the need to produce quarterly estimates, the panel nature of CE interview data provides a rich source of information to better understand patterns of nonresponse and potential nonresponse bias. Such information can also be used to target efforts to minimize nonresponse error in responsive design strategies (Olson, 2011).

The Diary survey also incorporates a panel design that interacts with the issue of attrition. Each selected household is asked to complete two waves of data collection, each wave being a one-week diary. Even though the waves are adjacent weeks from the same households, the estimation process considers each wave as an independent collection. A general issue with diary surveys is that compliance tends to deteriorate as the diary period progresses, that is, there is high compliance at the beginning of the period and less toward the end. The panel has not seen specific research on this issue for the CE Diary, but it is likely that there is less compliance during the second wave than in the first wave. This may be particularly true in households in which both diaries are placed at the same time without an intervening visit from the field representative. Without adjustment during the estimation process, it is possible that the lower reported expenditures in wave 2 will bring down the overall level of expenditures reported from the Diary survey.

Disproportionate Nonresponse

Having a high proportion of the initially sampled individuals respond to the survey may lead to higher quality data; however, it is neither a sufficient nor necessary condition. King et al. (2009) reported on four studies to examine potential nonresponse bias in the CE. One of these studies was discussed above. Even though the studies showed that the nonresponse was not missing completely at random (African Americans are underrepresented among the respondents, and those over 65 years old are overrepresented), they did not find measurable bias in estimating total expenditures due to the nonresponse.

A more recent study suggests a potential bias related to underrepresentation of the highest income households (Sabelhaus et al., 2011). The authors began by comparing CE estimates of income and the distribution of income with other relevant data sources such as the Current Population Survey, the Survey of Consumer Finances, and tax return–based datasets from the Statistics of Income. These comparisons show that the CE has relatively fewer completed surveys from households with income \$100,000 or greater. The authors also showed that the average income estimated per household for this highest income group is substantially below the estimated average from these other sources.¹⁰

The authors demonstrated by comparing the CE sample to geocoded tax data that higher income units are less likely to respond on the CE and are underrepresented even after weighting, while units at lower levels of income are adequately represented. While there is not a large difference in the total population counts, the underrepresentation of the upper income groups could lead to an undercount of income in the higher income levels and consequently also to understating the aggregate level of spending. They have concern because these high-income households may have a different mix of expenses when compared to other households. This differential could affect the relative budget shares calculated for the CPI. The authors speculate that a significant portion of the difference between CE aggregate spending and PCE spending might be accounted for by the nonresponse of higher income consumer units. They concluded that:

Only the very highest income households seem to be under-represented in the Consumer Expenditure Survey (CE), and the mystery of overall under-reported spending in the CE is not fully explained by that shortcoming. At least some of the shortfall in aggregate CE spending seems attributable

¹⁰Greenlees, Reece, and Zieschang (1982) carefully analyzed nonignorable income nonresponse (nonresponse related to the variable being imputed) using matched microdata from the CPS and IRS. They demonstrated clearly the problem that nonignorable nonresponse imposes.

by under-reported spending by at least some CE respondents. (Sabelhaus et al., 2012, p. 21)

Conclusion 5-15: Nonresponse is a continuing issue for the CE as it is for most household surveys. The panel nature of the CE is not sufficiently exploited for evaluating and correcting either for nonresponse bias in patterns of expenditure or for lower compliance in the second wave of the Diary survey. Nonresponse in the highest income group may be a major contributing factor to underestimates of expenditures.

ISSUES REGARDING NONEXPENDITURE DATA

The use of the CE data for economic research provides the impetus for collecting data on demographics, income, investment, and savings at the household level in the CE. The panel identified several issues in the current CE with these types of data that make the research process more difficult.

Reporting Periods for Income, Employment, and Expenditures

Ideally, researchers would like to have expenditures, income, and employment for responding households collected over the same reporting periods. The current Interview survey collects expenditure information for each of four quarters during a year. Income and employment information is collected for the previous 12 months, but only during the second and fifth interview. The current Diary survey collects expenditure data for two weeks, but income and employment data for the previous 12 months. The inconsistency of the collection periods for these different types of data can make it difficult to reconcile large differences in expenditure and income at the household level when these differences occur. While researchers have expressed the importance of having all data (including expenditures and income) collected over the same reference period, some panel members have expressed the opinion that it is also important to allow respondents to report for a period for which they can do so most accurately.

Demographics and Life Events

Examining the impact on household spending due to a variety of stimuli is important in the economic research done using the CE data. When changes occur, it is difficult to reconcile changes in expenditure and income without information about whether an individual household has undergone a major life event (e.g., marriage, divorce, or change in employment status) sometime during the year. The current CE collects relatively limited infor-

mation on these major life events. For example, the CE Interview survey does not collect changes in job status (and the reason for those changes) between survey waves.

Linking with Administrative Data Sources

The ability to link CE data to relevant administrative data sources (such as IRS data or data on program participation) could provide additional richness for economic research as well as providing potential avenues to investigate the impact of nonresponse on the survey results. Data confidentiality procedures have presented barriers to such linkage. Some success in this area has been achieved by some other federal surveys that ask respondents' permission to match their survey responses with administrative data. Some surveys have experimented with an "opt out" version, where respondents can say that they do not want the matching to occur (Pascale, 2011). These strategies would be useful to try for the CE.

Conclusion 5-16: For economic analyses, data on income, saving, and employment status are important to be collected on the CE along with expenditure data. Aligning these data over time periods, and collecting information on major life events of the household, will help researchers understand changes in income and expenditures of a household over time. Linkage of the CE data to relevant administrative data (such as the IRS and program participation) would provide additional richness, and possibly provide avenues to investigate the effect of nonresponse.

SUMMARY OF REASONS TO REDESIGN THE CE

This chapter specifically addresses the issues upon which the panel bases its recommendations in Chapter 6 to redesign the CE. The CE surveys appear to suffer from underreporting of expenditures. This conclusion is based on comparison of the CE estimates to several sources, but primarily to the PCE. The panel does not consider the PCE as "truth," but does consider the results informative. The comparisons were made considering categories of expenditures with comparable definitions between the CE and PCE. The overall pattern indicates that the estimates for larger items from the CE are closer to their comparable estimates from the PCE. The current Interview survey estimates these larger items more closely to the PCE than does the current Diary survey. For 36 smaller categories, neither the Interview survey nor the Diary survey consistently produces estimates that have a high ratio compared to the PCE. Thus, the panel concluded that there are underreporting issues with both the CE Diary and CE Interview surveys and

proceeded to review response and nonresponse issues that could contribute to this underreporting.

Before examining sources of potential response errors in the Interview survey and Diary survey separately, the panel observed that the mode, questions, and context used in the Interview survey and Diary survey are quite different. Therefore, one ought to expect differences, both in issues that need to be addressed and in the estimates obtained. It is therefore not surprising that different amounts are reported in the Interview and Diary in this situation, and that these differences are not always in the same direction.

The panel examined potential sources of response error in both surveys. They concluded that both the Interview and Diary surveys have issues with:

- motivation of respondents to report accurately,
- structure of data collection instruments that leads to reporting problems,
- recall or reporting period, and
- proxy reporting.

Additionally, they expressed concern about the infrequent use of records in the Interview survey that is less relevant to a concurrent mode of collection.

The Interview and Diary surveys have similar response rates of 73 to 72 percent. These rates are lower than for some important federal surveys, but appear to be better than consumer expenditure surveys in some other western countries. There is concern about attrition within the panel designs for both surveys, as well as concern about the effect of disproportionate nonresponse from the segment of the population in the highest income groups.

In sum, there are response and nonresponse issues with both the concurrent (Diary) and recall (Interview) collection of data in the CE as currently implemented. The panel does not conclude that one method is intrinsically better or worse than the other. However, it does believe that different approaches to these methods have the potential to mitigate these problems.

6

Pathway to an Improved Survey

This final chapter is based on the information and analysis in previous chapters of the report. It describes three potential prototypes to consider in redesigning the Consumer Expenditure Surveys (CE) and offers recommendations to the Bureau of Labor Statistics (BLS) from the Panel on Redesigning the BLS Consumer Expenditure Surveys on research and other inputs needed in redesigning the CE.

OVERVIEW

The CE has many purposes and a diverse set of data users. This is both the strength of the program and the foundation of its problems. The CE program tries to be “all things to all users.” The current design creates an undesirable level of burden on households and quality issues with its data. The Interview survey asks respondents for a very high level of detail collected over an entire year, with potentially counterproductive effects on their motivation and/or ability to report accurately. Because the Interview survey had been deemed not to satisfy all user needs, the program also includes the Diary survey. The Diary survey supplies much of the same information but at an even higher level of detail over a shorter period of time by using a different collection mode and a different set of respondents. However, Diary respondents appear to lack motivation to report consistently throughout the two-week collection period. Unfortunately, these two surveys are designed independently so the resulting data are not statistically consolidated to achieve their potential precision and usefulness.

The Consumer Price Index (CPI) drives the level of detail asked in the

current CE surveys. The CPI is a Principal Economic Indicator of the United States and a crucial user of the CE. The CPI currently uses CE data for over 800 different expenditure items to create the budget shares required for those indexes. Most, but not all, budget shares come from the CE. In theory, a number of survey designs can provide the information required by the CPI, collecting a significant level of expenditure data without inflicting the level of burden on households that the current CE does. These designs, including a number of “matrix type” sample designs, involve asking each household only a portion of the total detail required while using weighting and more sophisticated modeling to produce the needed estimates. The data from these types of designs can provide the needed level of detail with appropriate precision needed by the CPI but with less burden on each household (Eltinge and Gonzalez, 2007; Gonzales and Eltinge, 2008, 2009). This family of designs would also meet most of the needs of government agencies in program administration and would allow BLS to continue to publish standard expenditure data tables. However, these types of designs are not optimal for other uses of the CE.

Researchers and policy analysts use the CE microdata to examine the impact of policy changes on different groups of households and to study consumers’ spending habits and trends. Many such uses are described in Chapter 2. These data users generally do not need the same level of item level detail required by the CPI. To them, the value of the CE lies in the “complete picture” of demographics, expenditures, income, and assets all collected for the same household. A comprehensive set of data at the household level allows microdata users to look at the multivariate relationships between spending and income in different types of situations for different groups of households. These data users also use the “panel” component of the CE, which provides the same information for a given household over multiple quarters. Parker, Souleles, and Carroll (2011) and Chapter 4 (“Feedback from Data Users”) further describe the usefulness of panel data in this type of analysis.

The multiple and divergent CE data uses are difficult to satisfy efficiently within a single design. Survey designs always involve compromise, and the current CE design tries to provide the breadth and detail of data to meet the needs of all users and then compromises by accepting the heavy burden and unsatisfactory data quality that emerges. The panel recommends that BLS redesign the CE only *after* rethinking what those compromises should be so that the trade-offs associated with redesign possibilities can be articulated and assessed within a well-developed priority structure. Determining these types of priorities for the CE is ultimately the responsibility of BLS, and is beyond what would be appropriate or realistic for the panel to undertake. Therefore, the panel makes the following recommendation:

Recommendation 6-1: It is critical that BLS prioritize the many uses of the CE data so that it can make appropriate trade-offs as it considers redesign options. Improved data quality for data users and a reduction in burden for data providers should be very high on its priority list.

The panel recommends a major redesign of the CE once the priorities for a redesign are established. In its work, the panel concluded that many response and nonresponse issues in both the Diary and Interview surveys create burden and lead to quality problems with the expenditure data. The panel has concluded that less invasive cognitive and motivational corrections that might be made to improve recall and the reporting of specific expenditures would most likely increase overall burden. Since burden is inextricably connected with much of the survey's problems, increasing it would be counterproductive.

Recommendation 6-2: The panel recommends that BLS implement a major redesign of the CE. The cognitive and motivational issues associated with the current Diary and Interview surveys cannot be fixed through a series of minor changes.

The charge to this panel was to provide a “menu of comprehensive design options with the highest potential, not one specific all-or-nothing design” (see Appendix B). Before BLS sets prioritized objectives for the CE, the panel's most effective course of action is to suggest alternative design prototypes, each of which has a higher potential for success when enlisted to achieve a different prioritized set of objectives.

With that said, these prototypes share much common ground. The statistical independence of the current interview and diary samples is eliminated. The prototypes orient data collection methods toward an increasingly computer-literate society in which new tools can make the reporting tasks easier for respondents while providing more accurate data. The new prototypes are geared to increase the use of records and decrease the effects of proxy reporting. There is an increased emphasis on self-administration of survey components, while creating tools and an infrastructure that will monitor and support the respondent in these endeavors. The field representatives' role will still be important in directly collecting data, but their role will grow to also provide support in additional ways. The panel proposes incentives that will increase a respondent's motivation to comply and report accurately.

Finally and most importantly, all three prototypes propose new procedures and techniques that have not been researched, designed, and tested. The prototypes that the panel offers are contingent upon new research un-

dertakings and rigorous assessment. There is a lot of relevant background theory and research available, and the BLS research program and Gemini Project deserve praise for much of that work. However, the panel wishes to state clearly that the empirical evidence on how well each of the proposed prototypes would work is missing. As with the current CE surveys, the new prototypes include some diary-type data collection and some recall-type data collection. They include some self-administered data collection and some interviewer-assisted data collection. Notwithstanding, the new prototypes are sufficiently different from the current CE surveys that BLS cannot and should not use the current CE to extrapolate how well these prototypes will work in regard to response, accuracy, and underreporting. Considerable investment must be made in researching elements of the proposed designs, to find specific procedures that not only are workable but also are most effective. Some ideas will ultimately be successful, while others will be shown to have serious flaws. The critical point is that these prototypes are not operationally ready, and the process of selecting a prototype or components of a prototype for implementation should be based not only on BLS' prioritization of goals of the CE, but also on empirical evidence that the proposed procedures can meet those goals.

Recommendation 6-3: After a preliminary prioritization of goals of the new CE, the panel recommends that BLS fund two or three major feasibility studies to thoroughly investigate the performance of key aspects of the proposed designs. These studies will help provide the empirical basis for final decision making.

Issues related to nonexpenditure items on the CE were discussed in detail in Chapter 5. These issues include such things as synchronization of expenditure and nonexpenditure items over similar reference periods, and collecting changes in employment status and other life events. These types of issues are important to the research uses of the CE. The panel offers the following recommendation that should be viewed within the context of BLS prioritization of the goals of the CE.

Recommendation 6-4: A broader set of nonexpenditure items on the CE that are synchronized with expenditures will greatly improve the quality of data for research purposes, as well as the range of important issues that can be investigated with the data. The BLS should pay close attention to these issues in the redesign of the survey.

With a new design, some existing uses of data may fall by the wayside. New and important uses will emerge. BLS has a talented and knowledgeable staff of statisticians and researchers who have worked with the CE

for many years. They understand the survey well, and the cognitive issues described by the panel are not a surprise to that staff. Using the framework that the panel has put forward, BLS statisticians will be able to pull together and test the specific details of a redesign that is appropriate for BLS' collective priorities, budget, and time frame.

The rest of this chapter describes the three different prototypes, with many commonalities but each with its own focus. A more detailed discussion of those commonalities comes first, and then the report describes and compares the three prototypes. The final sections of the chapter begin a roadmap for moving toward a new design, including a discussion of important research issues.

PANEL'S APPROACH TO DESIGN AND THE COMMONALITIES THAT EMERGED

The panel considered many approaches to a redesign of the CE, and sorted through those numerous options by focusing on the following fundamentals:

- Improve data quality.
- Be mindful that the resources (both out-of-pocket and staff) available to support this survey are constrained.
- Be mindful that the survey processes have to be workable across the entire population of U.S. households—the more distinct processes that need to be designed for different population groups, the more resources will be required.
- Keep it simple—to the extent possible.
- Provide respondents with relief from the current burden level of the CE.
- Provide respondents with sufficient motivation to participate.
- Support the use of records and receipts.
- Support the current uses of the CE to the extent possible, and provide options in support of the prioritization of those uses in the future.
- Utilize newer data collection methodology and external data sources when supportive of the above fundamentals.

It is not reasonable for the panel to discuss all of the options that they considered and laid aside, but this section of the report is intended to illuminate the concepts and strategies that emerged with broad consensus during discussion of some of the major decision points in the panel's deliberations. These commonalities can be seen in the design of the three prototypes.

Implement a Major Integrated Redesign

The panel came to an early conclusion that the cognitive issues with the existing surveys cannot be fixed with minor corrections, and it would be a mistake to focus *independently* on the various cognitive issues addressed in this report. The best approach is an overall redesign of the CE, with component pieces being shaped to minimize these cognitive problems at each phase of data collection.

The sample design for the new CE should be developed with a view toward integrating sample panels and data collection periods on a panel via statistical modeling in the estimation process, rather than generating independent estimates for each panel and data collection period. This method ensures that all data collected within the design can be fully utilized to minimize variance of estimates by capitalizing on the temporal components of the design or by integrating sample panels that collect different, but related variables from respondents. It may be possible that sophisticated sample designs along with appropriate modeling can provide needed data products with reduced burden on respondents. In investigating this possibility, it will be important to avoid creating household-level data with such a complicated structure of measurement error or statistical dependencies that it makes research use very difficult. At least, any reductions in possible use of the data need to be consistent with newly clarified BLS priorities.

Reduce Burden

The extreme detail associated with the current CE, and the amount of time and effort it takes to report those details, are major causes of underreporting of expenditures. These need to be significantly reduced for most respondents. The panel identified a number of ways to reduce burden, and more than one burden-reducing concept is included in each redesign prototype. Burden-reducing opportunities include (1) reducing the overall detail that is collected on expenditures, income, and/or assets; (2) asking details (or certain sets of details) to only a subsample of respondents, providing burden relief for the remaining sample; (3) reducing the number of times a household is interviewed or the number of tasks they are asked to do; (4) reducing the overall sample size and using more sophisticated estimation and modeling to maintain levels of precision; and (5) making response cognitively and physically easier. The panel spent considerable time identifying ways to reduce burden. It realizes that several of these options may be at odds with collecting a complete picture of income and expenses from each individual household over longer reporting periods. This is why it is essential for BLS to further clarify its priorities for data uses, recognizing that one survey cannot satisfy all of the possible data users.

Use Incentives to Increase and Focus Motivation

The CE surveys are very complex and burdensome, and even with the burden-reducing changes, the CE will remain a difficult challenge for households. Respondents currently have little motivation to respond, or more precisely to respond accurately, on the CE. The panel anticipates that respondents will have additional responsibility under a redesign to keep and record expenditures. The panel collectively agreed that respondents needed greater motivation to carry out these tasks and proposed that an incentive structure composed of monetary and nonmonetary incentives should be developed and implemented. The structure should be based on the amount of effort asked of a respondent and used to effectively encourage record-keeping and reporting from those records. The panel speculated that the incentive payments would need to be fairly large to effect the needed motivation to report accurately. Components of an effective incentive program are discussed in more detail later in this chapter.

Support Accurate Use of Records

The panel envisions a redesign that will increase the respondents' use of records in reporting expenses. This can be accomplished in a variety of ways. In the three prototypes, incentives are offered. There is an increased emphasis in each prototype to incorporate supported self-administration in a way that provides a structure to promote accurate reporting and increased use of records. This means incorporating flexibility to allow respondents to provide data at a time and in a way that is most convenient for them, and to answer questions in the order that they prefer. It means redesign of data collection instruments (whether self-reports and interviewer-driven, paper or electronic), technology tools, training, reinforcement, and incentives to facilitate recordkeeping. Minimizing proxy reporting in the reporting of detailed information is another improvement that can lead to more accurate reporting and use of records and receipts.

Redesign Survey Instruments

The new CE will need to redesign data collection instruments so that they simplify the respondent's task. The panel sees a movement toward self-administered data collection with the field representative acting in a support role. However, the prototypes also incorporate interviewing by field representatives. Even though the panel envisions a wide acceptance of tablet-based interfaces, paper instruments will be needed for the foreseeable future. The current instruments may not suffice for this purpose.

Increase Use of Self-Administration

The panel discussed the advantages and disadvantages of various collection modes, and considered changes from the current CE surveys. The panel expressed concern about the shift in the current CE toward telephone data collection (primarily due to constrained resources), and felt this was not the best shift for data quality. The panel's final recommendations move toward self-administration of these complex surveys. There are several reasons for this shift. The first is to encourage the use of records as discussed in the paragraphs above. This mode allows respondents to provide data in a way and at a time that is most convenient for them. When paired with an appropriate incentive structure, it can encourage respondents to take the time needed to use those records and receipts. A second reason is to take advantage of newer technology that can allow consistent, remote monitoring of self-administered data collection without the cost of having an interviewer present.

Reduce Proxy Reporting

The current CE surveys use proxy reporting because of the additional cost associated with working separately with multiple survey respondents within a household. The panel looked for solutions that will allow (and encourage) individual members of a household to report their expenditures without the accompanying increase in cost. The solution is a shared household tablet that each member of the household can use to enter expenses, but there is still a "primary household respondent" who oversees the entire process. This solution does not provide confidential reporting, and thus does not solve the problem when household members are reluctant to share details of certain expenditures with other household members. However, it does have the potential to eliminate much of the current proxy report process with minimal added cost per household.

Utilize Newer Data Collection Technology

The time is right to emphasize new technological tools in data collection. This is an essential component of the panel's concept of supported self-administration. The panel discussed many technological alternatives and found one tool that was particularly appealing to the panel across a variety of designs—the tablet computer. The panel proposes the use of tablets in each of its redesign prototypes as an effective data collection tool. Lightweight and easy-to-use tablets represent stable (robust) technology, are commonplace, feature more than sufficient computing power, and are economical in price. The panel envisions that the tablet would sit on the

kitchen counter and be used by multiple household members in a “shared” approach to recording expenditures.

The panel also considered such alternatives as Web-based data collection, smart phone apps, and portable scanners for receipts. All are interesting tools and potentially could be used together in a redesigned CE. However, the panel stuck with its fundamentals—keep it simple and be mindful that the survey processes have to be workable across the entire population of U.S. households and that each additional approach (tool) will require additional resources to build and support. The panel looked for the one tool with the most potential.

Web data collection is not that tool. The Bureau of the Census (2010) estimated that only 44 percent of all U.S. households had Internet access either within or outside the home. This percentage varied greatly by household demographics and income. So requiring Internet access to use the electronic instrument would relegate the majority of households to the “paper” option. Additionally, building high quality Web-based instruments that work on multiple platforms (different computers, different browsers, high-speed versus dial-up Internet access, smart phone browsers) can be very resource intensive. By providing the tablet to the household, BLS would be developing for a single platform, and the panel hypothesizes that a substantially greater percentage of households will be able to use the tool than if BLS relied on Web collection.

The panel saw similar issues with using smart phone apps—lack of coverage of the population of households, and considerable variability in hardware and software platforms. These devices are growing in popularity, but BLS would have to develop and maintain multiple versions even for use within the same household.

Portable scanners would allow respondents to scan receipts and upload them to a waiting file. These devices could be used along with the tablet PC for use in recording receipts. However, the array of formats and abbreviations that are used on printed receipts would likely require considerable intervention after the scanning to properly record each expenditure. Adding these scanners to the household would also require additional training.

The use of technology tools, and the tablet PC in particular, is discussed in more detail later in this chapter. The panel will recommend that BLS begin using this one simple tool, knowing that its implementation will be challenge enough for the short run.

Use Administrative Data Appropriately but with Caution

The potential use of external records or alternative data sources as a replacement or adjunct to current survey data for the CE is often raised in discussions of a CE redesign. Whether at the aggregate or the micro level,

the appeal of “readily available” information appears, at first glance, to be low-hanging fruit. Although such information might hold great promise, upon closer inspection the panel also realized that use of these data is accompanied by increased risk and significant resource outlays. There is a cost/quality/risk trade-off that needs to be fully investigated and understood.

The panel discussed the potential use of these external data at the micro level and identified several concerns: Permission from household members to access such things as personal financial data, utility bills, and shopping data (loyalty card) would be difficult to obtain and thus replace only a small percentage of survey data; BLS would have to develop an in-house infrastructure to access and process data from each external source (this would be a significant drain on available resources); and BLS would have to continue to field a complete survey for the majority of households. That said, there are scenarios under which these data could be quite useful, particularly at a more macro level. However, caution is warranted. This subject is discussed in greater detail later in this chapter.

Create a Panel Component and Measure Life Event Changes

Economic analysts utilize the panel component of the current CE in much of their research. The report incorporates a panel component (with data collection from the same households at a minimum of two points in time) within each of the three prototypes. Each design also includes a re-measurement of income and “life events” (such as employment status, marital status, and disability) at each wave. However, the panel components differ considerably from design to design in the length of the response period, and this will significantly affect their relative usefulness in economic research. Of the three prototypes described, Design B has the most comprehensive panel component, with three waves and a response period of six months for each wave. Design C has two waves with a response period of three months for each wave. Design A has two waves, but with more variable response periods for each wave.

REDESIGN PROTOTYPES

In this section, the panel presents three specific redesign prototypes. All three designs meet the basic requirements presented in *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011). All three prototypes strive for increased use of records, incorporate self-administration (supported by the field representative, a tablet computer, and a centralized support facility) as a mode of data collection, and use incentives to motivate respondents. All three prototypes continue to use field representatives for interviewing and other support, and they all feature

either a single sample or integrated samples. However, each prototype is different—a better fit for a specific use of the data. BLS needs to prioritize the various data requirements of the CE and move toward a redesign that is best for its carefully considered prioritization. In overview,

- *Design A* focuses on obtaining expenditure data at a detailed level. To do this, the panel proposes a design with concurrent collection of expenditures through a “supported journal”—diary-type self-administered data collection with tools that reduce the effort of recordkeeping while encouraging the entry of expenditures when memory is fresh and receipts available. It also features a self-administered recall survey to collect larger and recurring expenses that need a longer reporting period. This design collects a complete picture of household expenses but with reports over different reporting periods.
- *Design B* provides expenditure data for 96 expenditure categories, rather than the more detailed expenses provided by Design A, but provides a complete picture of household expenditures over an 18-month period. It builds a dataset that would be excellent for economic and policy analysis. This design makes use of a recall interview coupled with a short supported journal. Two subsequent contacts with the same households are made over 13 months, repeating the original data collection using supported self-administration to the extent possible. This design also recommends that a small subsample be subsequently interviewed intensively over the following two calendar years, with collation of records, the use of financial software, and emphasis on a budget balance. This is discussed separately at the end of the description of Design B.
- *Design C* incorporates elements of both Designs A and B. It collects the detail of expense items as in Design A, while providing a household profile for six months. To do both, it uses a more complex sample design, collects different information from different samples, and requires more extensive use of modeling to provide expenditure estimates and the household profile.

Design A—Detailed Expenditures Through Self-Administration

This prototype features a sample of households with two data collection waves, each of which features the concurrent reporting of expenditures over a two-week period using a supported journal. The design also incorporates a self-administered recall survey for larger, less frequent expenses. The design maximizes the use of supported self-administration and concurrent reporting of expenses. Figure 6-1 provides a flow outline of Design A.

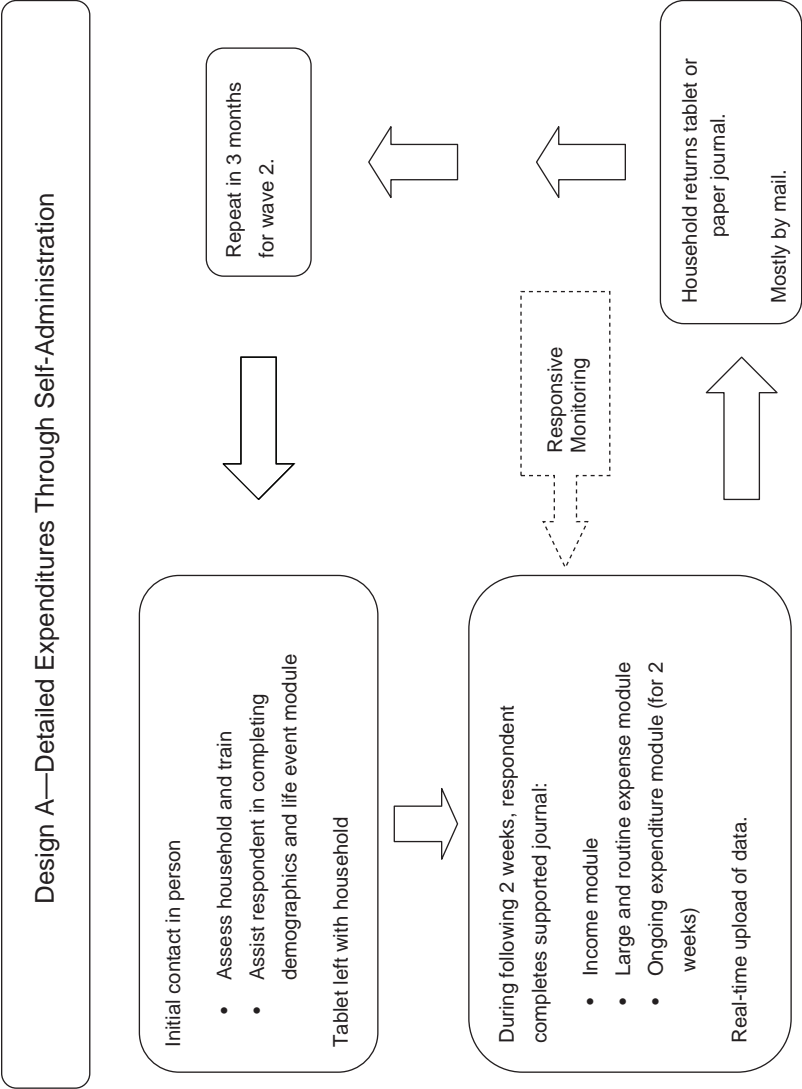


FIGURE 6-1 Process flow for Design A—Detailed Expenditures Through Self-Administration.

Objectives

The objective in Design A is to maximize the benefits that can be derived from self-administration in a new era of effective tablet computing and modular interface design. The idea is to simplify the respondent's task, to allow respondents to provide data at a time and in a way that is most convenient for them, and to answer questions in the order that they prefer. In doing so, the panel believes that the survey can collect detailed expenditures accurately with reduced burden on respondents. The goals are to

- promote accurate reporting of detailed expenditure data by allowing sufficient time and space for careful enumeration of expenditures while using records and receipts;
- reduce the effort it takes to report those expenditures by providing support and technology tools; and
- reduce respondents' tendencies (often implicitly encouraged in current methods) to estimate, guess, satisfice, or underreport.

Key Assumptions

Design A makes several key assumptions about the collection of consumer expenditure data:

- Detailed data for many items can best be obtained by supported concurrent self-administration, and a tablet-type device can assist in keeping the supported journal for most households.
- Interfaces for tablet-based applications (apps) will follow the best and newest principles of app design and testing, rather than simply importing or modifying current computer-assisted self-interview (CASI) technology and software.
- There will be a low refusal rate for use of the tablet given effective software and interface implementation.
- A two-week tablet-based reporting period for recording ongoing purchases is a plausible period. Households will be willing to participate in a second wave.
- For expenditures that must be recalled, different recall periods are appropriate for different categories of expenses.
- It is desirable to build alternative recall prompts or cues for different respondents who may have different ways of mentally organizing expenditures and/or different strategies for recalling them.
- The current set of income, asset, and demographic questions are reduced to only the essential items.
- Monetary incentives are available for respondents.

Sample Design

A single sample of households would be selected each quarter, with data collection initiated for a portion of the sample every two weeks within the quarter. A second wave of data collection for this same sample would take place in the subsequent quarter, with all responding households asked to participate for a second two-week collection period. Thus, for any given quarter, one independent sample would be initiated into wave 1 of the design, and a second sample would be brought forward from the previous quarter for wave 2 data collection.

Data Collection: Modes and Procedures

The predominant mode used in Design A is self-administered using a tablet PC. The data collection interface on the tablet would be modular, with flexibility on entering information in any desired order, with four modules:

- *Demographics and Life Events Module*: demographics and other information about the household. For wave 2, the questions would be modified to ask about key “life events” that might have transpired over the previous quarter.
- *Ongoing Expenditures Module*: for recording ongoing expenditures by household members during the two-week period.
- *Large and Routine Expense Module*: for reporting larger and routine expenses that may not be effectively measured during a two-week collection period.
- *Income Module*: household income, assets, and “labor force status” questions.

Navigation of the interface across and within the different modules would be streamlined and transparent for users of a wide range of backgrounds. The model for the interface would be “TurboTax” (commercially available tax preparation software) rather than the linear flow methodologies of today’s self-administered questionnaires. This means that the app would be built in modules and the user could choose to fill in the information in any order that is convenient. The user could also come back to any item and make a change at any time. Alternatively, the user (respondent) could choose to use a structured interview approach for moving step-by-step through the app.

During the initial in-person contact with the household, the field representative would identify the main “household respondent” and assist him or her with completing the *Demographics and Life Events Module* on the

tablet. The content of this first instrument would include standard demographics plus other items to assist in tailoring data collection instruments and estimation in the event that the household drops out of the panel, such as the number and age of household members, income bracket, purchasing habits, and use of online payments.

The first meeting also allows the field representative to assess whether the respondent (1) can be fully self-sufficient in using the tablet for later data collection, (2) will need additional support and monitoring during the next two weeks, or (3) would be likely to have severe difficulty with or refuse to use the tablet. The goal is to encourage the use of the tablet as much as possible while maintaining data quality.

Proxy reporting would be reduced by encouraging household members to record their own day-to-day expenditures in the *Ongoing Expenditures Module* during the two-week reporting period. The household respondent would guide the other household members in using this simple module. There would be no separate identification numbers to compartmentalize each member's entries.

The household respondent would be asked to complete the *Large and Routine Expense Module* and the *Income Module* at his/her convenience during the two-week period. This allows the household respondent the time to review the questions and gather records. The *Large and Routine Expense Module* would ask about major purchases or periodic expenses such as automobiles, appliances, and college tuition. This module would also ask about routine expenses such as utility bills, mortgage payments, and health insurance payments. Respondents would be given alternative ways of entering the data that reflect reporting periods that correspond to their records. The *Income Module* would ask basic questions about household income and assets with a recall period that is most convenient for the respondent to report accurately. It would include information about changes in assets over the year.

At the end of the two-week period, respondents mail the tablet back. In some cases, a field representative visit may be needed to ensure the return of the tablet and capture key missing data. The *Demographics and Life Events Module*, the *Large and Routine Expense Module*, and the *Income Module* would be appropriately modified for a household's second wave of data collection.

Frequency and Length of Measurement

Design A features two waves of data collection, each two weeks long, one quarter apart. All households are included in both waves of the panel. At a convenient time during the two-week reporting period, respondents

report larger and routine expenditures and income and assets via self-administered modules.

Recall Period

For the *Ongoing Expenditures Module*, household members enter purchases during the two-week period. Ideally respondents will enter data every day or almost every day. Different members of the household would be able to record their own expenditures. The centralized facility would be able to monitor and intervene if households do not enter data regularly or if there is evidence of poor data quality.

For the *Large and Routine Expense Module*, the recall period would vary between annual, quarterly, and monthly for different domains of expenditure, depending on which period has been shown to lead to the easiest and most accurate reporting. Within a single domain of expenditure, respondents may be given alternative ways of entering the data that reflect reporting periods that are easiest to provide based on the records to which they have access. In other words, the burden of recalculating amounts or recall periods for the needs of the survey is placed in the tablet software or centralized processing systems, and not the respondent.

The recall period for income and assets questions would be set to minimize measurement error and support estimation needs. This period is described as that which is most convenient for the respondent to report accurately while being close to the reporting period for expenditures.

Incentives

A guideline for incentive use is presented further in this chapter. In Design A, the panel estimates that \$200 in incentives would be required per household (\$100 for each two-week data collection period).

Role of Field Representative

The field representative's role changes radically in Design A, from being the prime interviewer and data recorder to being the facilitator who encourages, trains, and monitors in support of the respondent's thoughtful and accurate data entry. In this shift, the design attempts to change the burden and motivational structure of the current methods of collecting expenditure data. This entire process is described in more detail in "Using the Tablet PC for Self-Administered Data Collection" on p. 157. The role of the field representative will vary for households that use the paper-supported journal, and also across different tablet households depending

on their comfort with the technology, willingness to use records, and household composition.

Role or Expectations of the Respondent

From the respondent's perspective, Design A makes it possible for the survey instrument to be tailored to his/her particular needs: expenditure context, comfort with technology, access to and format of records, preferences in reporting format, and preferred recall cues.

The design proposes a modern modular interface that has the simplicity, guidance, and intuitiveness of the TurboTax interface, rather than adhering to constrained traditional survey interfaces most often used today. This means that the app would be built in modules, and the user could choose to fill in the information in any order that is convenient. The user can also come back to any item and make a change at any time. Alternatively, the user (respondent) could choose to use a structured interview approach for moving step-by-step through the app.

The tailoring includes a paper alternative, with new forms of support for the respondent, if necessary. The design also gives respondents the opportunity to get 24/7 real-time support, exactly when they need it, as they record their expenditures and answer recall questions.

In this design, the respondent takes on the primary role of providing the most accurate data possible, supported by records to as great extent as possible. What differs from the respondent's current role in the interview is that the respondent (after the initial meeting with the field representative) controls the time, pace, and order of data recording. In this sense, the design implicitly supports the idea that providing accurate data will take time and thoughtfulness. What differs from the respondent's current role in the diary is that the respondent is provided an easy-to-use entry device augmented by ongoing interactive encouragement and support from both the interface and from remote support staff.

While in one sense, the respondent in the proposed design has greater responsibility (and support) than required in current expenditure surveys, in another sense the locus of responsibility is more distributed than before: among the respondent, field representative, and remote monitors (and, in a way, the interface designers and researchers). The respondent's role thus includes being aware of and taking advantage of as much remote or in-person support as is needed to allow accurate data reporting.

Post-Data Collection Analysis

Data collected through the tablet (demographics, expenditures, and income/assets) would be complete and in an appropriate format for pro-

cessing by the end of the data collection period. Monitoring of the reported data allows for ongoing edit-checking during the two-week reporting period. The interface within the tablet converts data entered in the *Large and Routine Expense Module* and the *Income Module* to appropriate standardized reporting periods.

For paper households, data entry by the field representative or central staff would be required. Depending on the state of the paper materials, this could be very simple data entry or could include more complex decision-making from an envelope of saved receipts. Future efforts could create a smart data repository of the sort envisioned in the Westat proposal to reduce human effort of this sort, but the panel does not propose that here.

Infrastructure

This design would require

- purchase and inventory control/maintenance of tablet PCs for use in the field;
- new resources for app development and database management, not only for initial design efforts but also for ongoing continuing development and management. At least initially, BLS may want to consider outsourcing this function to organizations with experience in app development and management;
- new training to support field representatives in their new roles (and possibly new hiring practices to attract field representatives with different backgrounds or skills);
- new centralized survey management infrastructure for monitoring and support: case-flow management, tracking progress, managing interventions, providing positive reinforcements and managing incentives. This includes appropriate staff for technical support as well as staff who can support respondents remotely and staffing a 24/7 help desk that respondents could access by pushing a button on the tablet;
- fewer field visits requiring fewer field representatives in the field; and
- ongoing research to implement this prototype and to keep abreast of changes and future directions in technologies and technology adoption and of how these affect respondents' recordkeeping and reporting proclivities.

Sample Size and Cost

The sample sizes and costs (data collection only) for Design A presented in this section are broad estimates for use only to compare across alternatives. More careful calculations were beyond the information and resources available to this panel. These calculations are based in part on a spreadsheet of 2010 CE costs provided to the panel. The panel used those costs to estimate for similar activities, and then calculated an estimated cost per sample case for the new prototype. Thus, these costs are for data collection only, and for data collection within a mature process. Estimates of sample size (net of 25% nonresponse) were then made that would keep to a neutral budget.

Assumptions:

- Cost per supported journal placement—\$165.
- 85 percent of households would use tablet. Costs would go down with greater use of tablet.
- Cost of tablet—\$200. It could be used 6 times, with an expected loss of 10 percent.
- Remote monitoring of responses—\$100 per two-week supported journal.
- In-person monitoring—\$150. Necessary for 10 percent of tablet households, and twice per each paper household.
- Incentives—\$100 per two-week supported journal.
- Paper processing for paper households—\$100.

Based on these assumptions, and remaining budget neutral (\$16,000,000), the panel calculated the following projections for this prototype:

- Total cost = \$16,000,000.
- Annual effective sample size (assuming 75% response) = 18,700.
- Average cost per sample = \$853.

Meeting CE Requirements and Redesign Goals

This prototype meets the basic CE requirements laid out in *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011). Additionally, it is designed to reduce burden, reduce underreporting of expenditures, and utilize a proactive data collection mode with newer technology. Table 6-1 provides greater detail.

TABLE 6-1 Meeting the CE Requirements and Redesign Goals with Design A

Goal	Design A—Detailed Expenditures Through Self-Administration
Produce quarterly and annual estimates of 96 categories of expense items	The design collects data at a more detailed level of expenditures than the 96 categories. Collectively, the 26 two-week data collections over the year will provide annual and quarterly estimates of expenditures at a fairly detailed level.
Income estimated over the same time period	Income is reported for a period most convenient for the respondent to report accurately while being close to the reporting period for expenditures. Income, like expenditure data, may be modeled for the entire year. Income, assets, and “labor force status” are requested during both waves.
Complete picture for household spending	A complete picture of each household is collected for two two-week periods, as well as information on larger items and routine expenditures over a longer recall period. By developing a process of seasonal adjustment for the four weeks of expense reporting, it would be possible to make quarterly or annual estimates of expenditures at the household level. The accuracy of those estimates will need to be researched.
Proactive data collection rather than change at the margins	The focus of this prototype is proactive self-reporting. Some larger expenditures are recalled, but in a setting to encourage the respondent to think about the expenditures and look up records.
Panel component with at least two contacts	There are two contacts for each household, in adjacent quarters.
Reduced burden	The proposed design redistributes burden in current Interview and Diary methods to a supported journal with a modern tablet interface. Burden is reduced by making the response tasks easier and more intuitive. This prototype reduces the number of contacts per household. Incentives are used to reduce perceived burden.
Reduced underreporting	Focus on detail, the increased use of records, allowing the respondent to record expenses at a time best suited and in a way best suited to him/her, reduced proxy responding, and incentives to perform the task are expected to lead to a reduction in underreporting of expenditures.
Budget neutral	Sample sizes are calculated to maintain approximately the current budget level.

Targeted Research Required

Many of the research requirements for Design A are common to the other two prototypes, and are discussed in more detail in “Targeted Research Needed for Successful Implementation of Design Elements” on p. 178. Research specific to this prototype includes studies that would de-

velop models that would estimate quarterly and annual expenditures and income at the household level from the four weeks of reported detailed data from the *Ongoing Expenditures Module* plus the data reported on the *Large and Routine Expense Module*.

Design B—A Comprehensive Picture of Expenditures and Income

This prototype attempts to collect income and expenses at the individual household level over 18 months. It features household respondents “recalling” expenditure data aggregated for the 96 categories of expenditures (discussed in Henderson et al., 2011) for the previous six months. It is anticipated that a more focused questionnaire with less categorical and chronological detail may take less time to complete than the current Interview survey. Coupled with three contacts instead of five, there is an expected overall burden reduction compared to the current CE Interview survey. If an effective supported journal can be designed, the same households in the sample would also be asked to participate in a one-week supported journal to collect detail on smaller expenses used primarily to disaggregate some expenses reported in the recall survey.

The initial contact for Design B is an in-person visit by the field representative to the household. The field representative would assist the respondent in completing a recall survey of expenditures on a tablet computer. The tablet would be left with the household for use in a one-week supported-journal concurrent collection of expenditures and then mailed back, similar to that described in Design A. The tablet would be returned to the household by mail in six months and again in one year to repeat the recall and supported journal in a self-administered mode. Figure 6-2 provides a flow outline of Design B.

Objectives

Design B has the following primary objectives:

- Rely on a basic recall mode of reporting, which has provided aggregate expenditure estimates in the past that were more in line with the Personal Consumption Expenditures (PCE).
- Redesign the recall questionnaire to collect expenditures directly from the respondent at a broader level of aggregation, rather than collecting the current level of detail and then calculating aggregates.
- Provide specific instructions to help the respondent estimate expenditures that cannot be recalled.

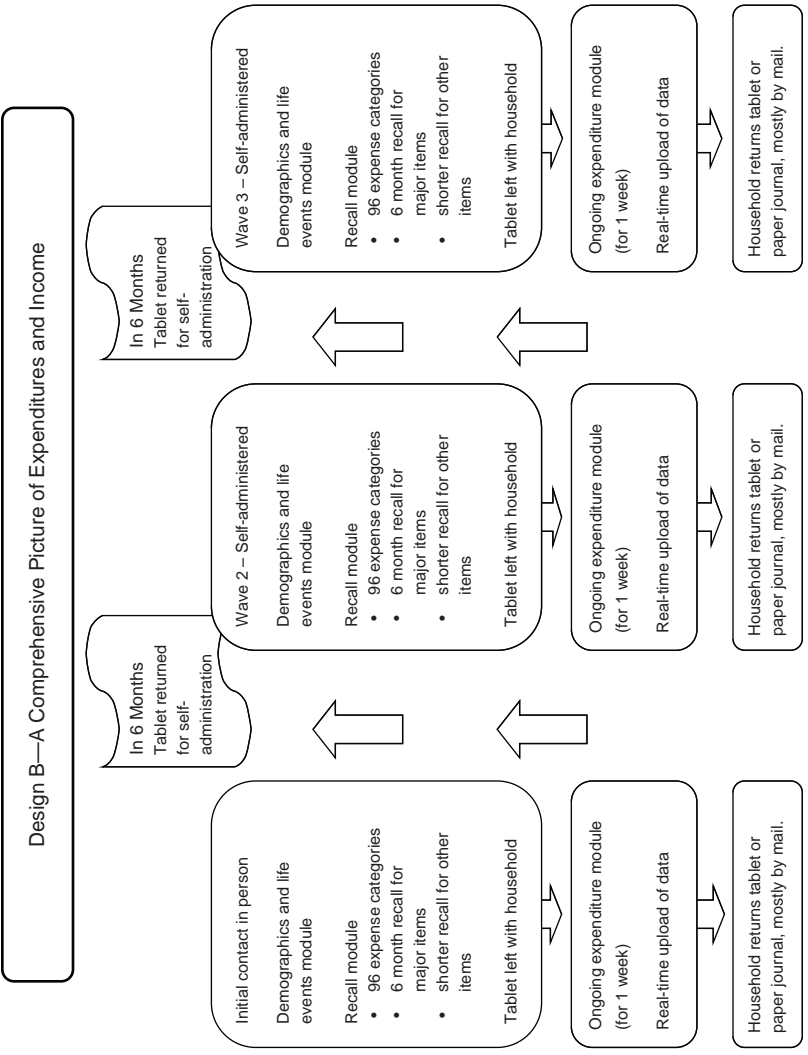


FIGURE 6-2 Process flow for Design B—A Comprehensive Picture of Expenditures and Income.

- Utilize technology to assist and support the respondent in filling in this redesigned questionnaire through a self-administered process in waves 2 and 3.
- Build data files well designed for economic research and policy analysis by providing a comprehensive picture of expenditure, income, and assets for each household for 18 months.
- Use a component subsample linked to the main sample that could be used to explore the accuracy of the overall data collection and provide an opportunity to collect data for more demanding research needs. It would employ techniques such as the prior collation of records, the use of financial software, and budget balancing. This component is discussed separately at the end of the description of Design B.

Key Assumptions

Design B makes several key assumptions about the collection of consumer expenditure data:

- The collection of “bounding” data is unnecessary and inefficient for the collection of accurate expenditure data in a recall survey.
- The value of the micro dataset flows primarily from the construction of expenditure data for relatively broad aggregates—for instance, the 96 expenditure categories for which CE tables are currently published—rather than an extremely detailed breakdown of those expenditures.
- A household can accurately recall aggregated expenditures for periods up to six months for some categories of expenditures. For other expenditure categories, respondents can approximate averages (e.g., average monthly spending for gasoline) that can be used to construct a full set of microdata for the entire six-month period.
- Household respondents will agree to remain in the panel for 13 months, with three data collection events during that period.
- The use of supported self-administration with the tablet, a central support facility, and the field representative allows most respondents to complete the redesigned recall questionnaire without the need for an in-person interview during waves 2 and 3.
- Households selected for the intensive subsample would be willing to participate in this more intensive data collection and are able, with assistance, to budget balance their annual finances.

To the extent these assumptions are valid, detailed microdata suitable for socioeconomic analysis can be obtained at the same time that the aggregate data objectives of the survey are met.

Sample Design

In Design B, a large sample of households is surveyed three times, at six-month intervals. To smooth the operation of the survey, one-twelfth of the households would be initiated each month. Once fully implemented, the workload in each month would be the initiation of a new survey panel, administration of the second wave of the survey to the households that had been initiated six months earlier, and administration of the final wave of the survey to the households initiated one year earlier. The field representatives are used intensively in the initial wave. If self-administered collection methods are successful, the field representative is used in the two additional waves only for households who need in-person assistance in completing the questionnaire.

In addition to a recall of expenses, each household would be asked to keep a one-week supported journal for the upcoming week. Thus, both recall and supported journal surveys are conducted on the same households, and repeated at six-month intervals for three repetitions.

Data Collection: Modes and Procedures

Design B collects recalled expenditure data, as well as demographic information, income, and assets. The first wave of collection is interviewer assisted, with the subsequent two waves relying on supported self-administration to the extent reasonable. As with Design A, survey instruments are presented through an interface on a tablet computer. The tablet interface follows the guidelines described under Design A and in “Using the Tablet Computer for Self-Administered Data Collection” on p. 157. Thus, the details are not repeated here.

In this prototype, the tablet is set up with three modules all designed for self-administration:

- *Demographics and Life Events Module*: demographics and other information about the household;
- *Recall Module*: for reporting expenditure, income, and asset data recalled or estimated for the past six months; and
- *Ongoing Expenditures Module*: for recording detailed, ongoing expenditures by household members during the one-week period of supported journal collection.

During the initial in-person contact with the household, the field representative identifies the main “household respondent.” As with Design A, the field representative assists him or her with completing the *Demographics and Life Events Module* on the tablet. The first meeting also allows the field representative to assess whether the respondent can be fully self-sufficient in using the tablet for data collection. The goal is to encourage the use of the tablet as much as possible in subsequent waves while maintaining data quality.

The field representative would then assist the respondent in completing the *Recall Module* reporting expenditures and income for the previous six months. This process not only obtains the needed data for this wave, but also trains the respondent on using the tablet and the specific modules in preparation for waves 2 and 3.

Following the completion of these modules, the field representative would ask the respondent to keep a concurrent supported journal for the next week (*Ongoing Expenditures Module*). Following the supported journal week, the respondent returns the tablet by mail.

Wave 2 takes place in six months and wave 3 in one year following wave 1. Respondents would be contacted a month ahead of the recall data collection and reminded to gather records and receipts. The tablet with the same data collection modules is used. Household respondents who were successful in using the tablet during wave 1 would be mailed a tablet for waves 2 and 3 to be completed without interviewer assistance. The “interviewer assistance” during wave 1 prepares them to complete this task. They are asked to complete all three modules and return the tablet following the supported journal week. Households that were not successful using the tablet in wave 1 would be contacted in person or over the telephone to complete waves 2 and 3.

Demographics and Life Events Module: This module includes standard demographics plus other items to assist in tailoring estimation in the event that the household drops out of the panel. Examples include the number and age of household members, income bracket, purchasing habits, and use of online payments. In waves 2 and 3, the questions would be modified to ask about key “life events” that might have transpired over the previous six months.

Recall Module: This module collects data on expenditures at a relatively broad level of aggregation based on variable recall periods. The objective is to obtain six-month estimates for each expenditure category. The assumption in this prototype is that reporting certain expenses (such as major durable goods, rent, and utilities) for six months would be relatively easy and accurate. For other recurring expenses such as groceries and gasoline,

for which shorter recall periods are appropriate, the respondent could make estimates of monthly averages of sufficient accuracy to be expanded to estimate spending for six months. Ultimately, all reported data would be expanded to a six-month estimate by the tablet app. The instrument would also include special prompts for expenditure categories that have been historically underreported, such as clothing, food away from home, alcohol, and tobacco. These prompts would include the types of expenditures and the possible locations of potential purchases.

Compared to the current CE Interview survey, Design B significantly curtails the number of questions both in terms of the breadth of categories (as discussed above) and the detail required in reporting those expenditures. For instance, the instrument would not ask in which month purchases were made, or sales tax, or information about the individuals for whom the purchase was made. These details add significantly to the burden in the current CE Interview survey.

Income, assets, and “labor force status” would be collected for each household for the same six-month period. Again, some estimation may be necessary based on current pay stubs and prior tax records.

Ongoing Expenditures Module: After completing Design B’s recall survey, households would be asked to keep a one-week self-administered supported journal. The supported journal would collect information on expenditure categories for which recall collection is more problematic and would provide additional detail that can be used to disaggregate totals. Thus, the supported journal would focus on specific categories of expenses. (However, research might indicate that it is more efficient to collect a complete record.) A paper-supported journal would be available for use when necessary, but the tablet would be the preferred mode.

Frequency and Length of Measurement

There are three waves of data collection at six-month intervals in Design B. Each wave uses a tablet computer with three modules. Wave 1 requires an in-person visit by the field representative with an “assisted” interview using the tablet for the *Demographics and Life Events Module* and the *Recall Module*. The tablet is left with the household for a one-week supported journal. In subsequent waves, the tablet is mailed to the household for self-administration of two modules plus the one-week supported journal. A household is in the sample for 13 months and recalled expenditure data are collected for 18 months. All recall and supported journal instruments are the same for different waves. The demographics module would be modified for use in waves 2 and 3 to ask about key “life events” that might have transpired over the previous six months.

Completing the *Demographics and Life Events Module* and the *Recall Module* is expected to take 30–45 minutes regardless of mode. The *Ongoing Expenditures Module (supported journal)* is expected to take 20 minutes per day.

Recall Period

The recall period for Design B's *Recall Module* is six months. If appropriate, the recall period for some items can be less than six months, and the expenditures reported for the shorter period used to estimate spending for the full six-month period. For the *Ongoing Expenditures Module (supported journal)*, the collection period is one week. The supported journal collects data on a daily basis, with little or no recall required.

Incentives

The panel estimates that an incentive payment of \$100 would be made to households for completing each of three waves of data collection in Design B. Households that were unable to use the tablet and require in-person enumeration for each wave would receive an incentive of \$50 per wave instead of \$100.

Role of the Field Representative

In Design B, the field representative establishes contact with each household and secures cooperation. The field representative conducts the first interview by assisting the respondent with wave 1 data including household composition, demographics, and the six-month recall categories. Simultaneously, the field representative trains the respondent on the use of the tablet and its modules in preparation for waves 2 and 3. The field representative recruits the household into the supported journal collection, emphasizing the tablet-supported journal collection in the vast majority of cases. The field representative trains the respondent on the use of the tablet for supported journal collection. The field representative explains the incentive for filling in the supported journal, leaves the tablet and a mailer at the household, and departs.

After the initial in-person visit, ideally the field representative never visits the household again. In principle, the household faithfully fills in the *Ongoing Expenditures Module (supported journal)* and mails the tablet back in a timely manner, and the household respondent successfully masters using the modules on the tablet and agrees to complete waves 2 and 3 in a self-administered mode. The field representative may be called upon by the central office to intervene with a household, for example, to encourage

reporting on the supported journal for the entire week, to answer questions about the instrument, to substitute a paper-supported journal for the tablet-supported journal, or to pick up the supported journal. The field representative does not monitor the cases in the first instance, but intervenes only at prompting from the central office.

The role for the central office includes monitoring the cases on a daily basis, phoning the household to intervene if the supported journal is not filled in, and prompting field representatives if their assistance is needed. The central office also fills the role of first-line support for respondent questions, forgotten passwords, and the like. The central office will also re-initiate contact with the household in six months, mailing the tablet and instructions.

Role or Expectations of the Respondent

The respondent provides data three times in Design B. The first time, the respondent is assisted by the field representative for the recall categories, but then records concurrent expenditures with the supported journal items for a week. The respondent may need to use a paper-supported journal rather than the tablet. Whether the supported journal is a tablet or paper, the respondent is expected to mail it back to the home office. The assumption of Design B is that most respondents are willing to use the tablet throughout all three waves. The tablet is expected to make the respondent's job easier.

Post-Data Collection Analysis

The spending for some goods over the previous six months would have to be estimated based on the spending during a shorter recall period. This could be as simple as doubling the expenditures reported for a three-month recall period, as Statistics Canada does for their interview survey. It is expected that the tablet app will make these adjustments. This design makes it easier for analysis because any modeling required for a complete record is done within the household, not across households.

Infrastructure

Infrastructure needs for Design B are similar to those in Design A and not repeated here.

Sample Size and Cost for Basic Component

The sample sizes and costs (data collection only) presented in this section for Design B are broad estimates for use only to compare across alternatives. More careful calculations were beyond the information and resources available to this panel. These calculations are based in part on a spreadsheet of 2010 CE costs provided to the panel. The panel used those costs to estimate for similar activities, and then calculated an estimated cost per sample for the new prototype. Thus these costs are for data collection only, and for data collection within a mature process. Estimates of sample size (net of 25% nonresponse) were then made that would keep to a neutral budget. Sample size and costs for the subsampled component are provided separately.

Assumptions:

- Cost per in-person recall module or interview—\$325.
- 85 percent of households would use tablet. Costs go down with greater use of tablet.
- Paper households are contacted in person on each wave.
- Cost of tablet—\$200. It can be used six times, with an expected loss of 10 percent.
- Remote monitoring of responses—\$100 per wave per tablet household.
- In-person monitoring—\$150, or \$325 for full interview. Necessary for 10 percent of tablet households, and once or twice per paper household per wave.
- Incentives—\$100 per wave per tablet household and \$50 per wave when interview is required. Households participating in the intensive study would receive \$150 per wave.
- Paper processing for paper households—\$100 per wave.

Based on these assumptions, and remaining budget neutral (\$16,000,000), the panel calculated the following projections for this prototype: (Sample size and costs for the subsampled component are provided separately.)

- Total cost = \$13,900,000.
- Annual effective sample size (assuming 75% response) = 12,200.
- Average cost per sample = \$1,138.

Meeting CE Requirements and Redesign Goals

Design B meets the basic CE requirements laid out in *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011). Additionally it is designed to reduce burden and reducing underreporting of expenditures. Table 6-2 provides more detail.

TABLE 6-2 Meeting the CE Requirements and Redesign Goals with Design B

Goal	Design B—A Comprehensive Picture of Expenditures and Income
Produce quarterly and annual estimates of 96 categories of expense items	All 96 categories of expense data are collected with three data points for each.
Income estimated over the same time period	Questions on income, assets, and “labor force status” are asked on each wave and for the same reporting period.
Complete picture for household spending	This prototype focuses on providing an improved picture at the household level over the current CE. Some items are collected in the recall module and some in the supported journal, but both are collected from the same households. Data adjustment (expansion) to the entire six-month period will be needed for items not collected or estimated for the entire six-month period. However, each household will have complete (or estimated) records for all 96 items.
Proactive data collection rather than change at the margins	The supported journal collection is proactive. The recall module is recall.
Panel component with at least two contacts	There are three contacts for each panel member.
Reduced burden	There are three administrations per household, instead of five for the current interview. The interview is less burdensome than the current CE in terms of length and in difficulty of task. However, the household is now expected to execute both the interview and the supported journal. Incentives are used to reduce perceived burden.
Reduced underreporting	This prototype makes the assumption that aggregating expenditures for recall over a six-month period will have less underreporting than an attempt to recall more detailed expenses over three months. Research is needed to evaluate this assumption.
Budget neutral	Sample sizes are calculated to maintain approximately the current budget level.

Targeted Research Needs

Many of the research requirements for Design B are common to the other two prototypes, and are discussed in more detail in “Targeted Research Needed for Successful Implementation of Design Elements” on p. 178. Research specific to this prototype includes the following:

- investigate the assumption that a “bounding” interview is unnecessary to avoid telescoping and other issues;
- investigate the accuracy and completeness of aggregated expenditures for periods up to six months and for estimates of averages (e.g., average monthly spending for gasoline) used in this prototype to construct a full set of microdata for the entire six-month period;
- develop appropriate models to “disaggregate” aggregated expenses using data from the one-week supported journal; and
- develop successful methodology for a component that will use an intensive interview and process based on prior collation of records and financial software to achieve a budget balance for the year at the household level, as described below.

Intensive Subsample in Design B

Design Objectives: A relatively small subsample of households who have completed wave 3 of the basic component of Design B would be asked to participate in a more intensive process to provide a full picture of income and expenditures over two consecutive calendar years. The process uses paper and online records more intensively, encourages the use of financial planning software, and employs budget balancing to reduce discrepancies between expenditures and income net of savings.

As discussed in Chapter 5, the PCE is the primary but problematic benchmark for comparing the CE aggregate expenditures. This subsample would attempt to establish accurate measurements of expenditures, income, and assets at the household level for a year through a more intensive record- and budget balance-driven process. Besides establishing an improved benchmark for measuring the success of the data collection methodologies in the basic component, the subsample would inform how better to collect expenditures in the ongoing survey and measure the extent and organization of household recordkeeping.

Key Assumptions: Assumptions for Design B’s intensive subsample include the following:

- Households that have completed wave 3 of the basic component of this prototype are willing to participate in this more intensive process if selected for the subsample.
- Respondents, with the help of the field representative, can reach reasonable balance between expenditures and income less savings. Fricker, Kopp, and To (2011) found the actual balancing between income and expenses difficult in practice. However, Statistics Canada used this approach for a number of years before redesigning their survey in 2009, so there is likelihood that a workable prototype can be developed.

Sample Design: A relatively small subsample of households that have completed wave 3 of the basic component of this prototype would be selected for this component. Calculations in this report are based on a subsample of approximately 5 percent of the original sample.

Data Collection: Modes and Procedures: The initial wave of data collection would begin approximately two months following the wave 3 interview. A second wave of data collection would be one year later. Data collection is expected to be face-to-face. It may take multiple visits to achieve the required balance.

Expenditure and income data reported during waves 2 and 3 of the basic component would be available to the respondent and field representative to work together to bring things in balance. The goal of having a financial budget that balances would be explained up front to the household. The respondent would be encouraged to use financial software and supply records, and/or draw information from online financial sources, including credit card and bank accounts as is done by Mint.com (2011). Additionally, the respondent will be asked about loyalty card programs in which they participate, and will be asked to provide permission for BLS to obtain the household spending data captured within those programs. Categories of spending for which there are no records become the leftover or residual part of expenditures. The main survey instrument is then used to fill in the missing parts of expenditures over the year, keeping in mind the need for the budget to balance.

Frequency and Length of Measurement: This design features two waves of data collection, one year apart. All households in the subsample will be included in both waves. Income, savings, and expenditure data for one year are required for each wave.

Recall Period: This intensive process focuses on obtaining and using records of income and expenses. However, during the balancing process, respondents may be asked about expenses and income for the previous year.

Incentives: The panel understands that the effort required of a respondent to actively participate in the financial balancing activities is greater than for a recall interview. Therefore, it estimates that an incentive payment of \$150 would be made to households for completing each of two waves of data collection, for a total of \$300.

Role of the Field Representative: The field representative's role in this component changes radically. He or she assists the respondent in sorting through available records of income and expenses over the year. The field representative would also work with the respondent to balance the household financial budget for the year, probing for additional income and/or expenses to bring the budget into balance.

Role or Expectations of the Respondent: The respondent is expected to be an active participant in this process to balance the components of the household's financial budget for the year. This would include providing paper and electronic records, and giving permission to access credit card, banking, and tax records directly.

Post-Data Collection Analysis: N/A

Infrastructure: None specific for this component of Design B.

Sample Size and Cost: The sample size and costs (data collection only) presented here are broad estimates for use only to compare across alternatives. More careful calculations were beyond the information and resources available to this panel. These costs are in addition to the costs for the basic component of Design B.

Assumptions:

- Budget balancing process may take several interviews with the household and may require more experienced field representatives. Expected cost per household—\$800 per wave, with \$350 per refusal contact.
- Response rate—50 percent
- Incentives—\$150 per household per wave.

Based on these assumptions, the panel calculated the following projections for this prototype.

- Total completed sample size per wave = 800.
- Average cost per household per wave = \$1,300.
- Approximate total cost = \$2,100,000.

Design C—Dividing Tasks Among Multiple Integrated Samples

Design C utilizes a multiphase sampling design that empowers estimation and modeling to provide the needed data products with reduced burden on respondents. In doing so, it provides detailed expenditures similar to Design A. It provides a complete picture of household expenses and income as in Design B, but for six months (instead of 18) and only for a portion of households. It features supported self-administration using a tablet computer and data collection interfaces as described in the other two designs. Figure 6-3 provides a flow outline of Design C.

Design Objectives

The guiding principle behind Design C is to avoid asking every household in the sample to perform exactly the same tasks. By dividing up the tasks, the overall burden on an individual household is reduced. The totality of information is brought together through estimation and modeling. The power of this design relies on (1) achieving good correlation between the estimates from the base survey and the estimates from the later phases of data collection; (2) developing effective models involving covariates such as demographic characteristics to connect estimates from the different subsampled surveys; and (3) achieving improvements in the data quality and reporting rates in the supported self-administered procedures used with the subsamples. It also provides panel data and a complete picture of a household on a subset of the overall sample.

Key Assumptions

Design C makes several key assumptions about the collection of consumer expenditure data:

- All the key assumptions listed under Design A are present.
- Strong correlations exist for covariates measured in the base survey and later phases of data collection.
- Effective models involving covariates such as demographic characteristics can construct estimates using the different subsampled surveys.
- Collecting a complete picture of household expenses and income on a substantial component of the overall sample with modeling of

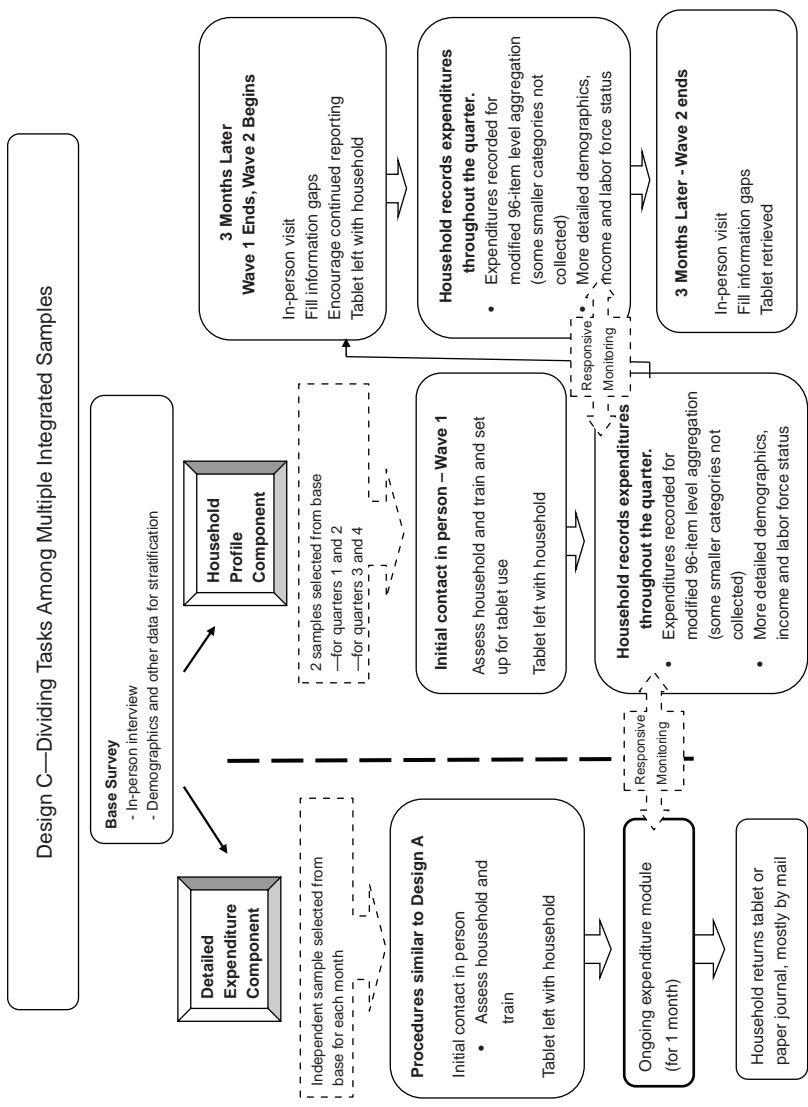


FIGURE 6-3 Process flow for Design C—Dividing Tasks Among Multiple Integrated Samples.

smaller expenditure items from a different sample will be sufficient for economic and policy analysis.

- A one-month tablet-based reporting period for documenting ongoing purchases is plausible. This is an open research question.

Sample Design

Design C includes a base survey followed by surveys of more intensive and frequent measurements. The base survey would have a relatively large sample size, collecting information to use for stratification and modeling. From that base, two sets of independent samples would be drawn. Sampled units in the first component are asked to keep a supported journal for one month to proactively record detailed expenditures. Sampled units in the second component are contacted for quarterly recording of aggregate expenses for two quarters.

The Base Survey: The base sample is a large, address-based sample similar to the current CE Interview sample. This initial survey forms a stratification base for sampling later phases of more intensive data collection. Base survey data are also used in models, combining them with data from later phases to produce estimates. In order to keep the base sample “fresh,” it would be supplemented each quarter with new samples that would be interviewed. This allows base survey data collection to go on throughout the year, and the samples for more intensive data collection to be selected using an updated base.

Detailed Expenditure Component: Design C calls for selecting 12 independent samples (one for each month) from the base survey and asking households in those samples to keep a supported journal of expenditures for one month. The purpose of these surveys is to proactively collect detailed expenditure data in the same manner as described in Design A. These data would be used in national and regional estimates, with precision enhanced by modeling back to the base survey and combining with data collected from the household profile component.

Household Profile Component: Design C calls for a separate component of the overall sample to focus on providing a complete profile of household expenses and income over two consecutive three-month periods. Independent quarterly samples would be selected from the base survey for collection of data through a combined proactive/recall collection process.

Data Collection: Mode and Procedures

The Base Survey: The base survey would be conducted in person using a computer-assisted personal interviewing (CAPI) instrument, administered on a “rolling basis” to provide the samples of the component surveys for the upcoming quarter. The survey would collect household demographics and basic income and asset data as needed for stratification, modeling, and imputation. (Perhaps this is as simple as a “range” value for household income, along with presence or absence of various assets.) It would collect global expenditures, at the 96-item level, aggregated even more if the correlations allow, with a variable recall period based on expense item. There would also be a series of questions regarding the household’s basic expenditure habits and comfort level with various reporting technologies.

Detailed Expenditure Component: The *detailed expenditure component* of Design C is set up in the same way as the supported self-administered journal described in Design A. One difference is that, in this prototype, households are asked to maintain the supported journal for four consecutive weeks. In Design A, households were asked to maintain the supported journal for two weeks, and then asked again to complete the two-week supported journal the following quarter.

Another difference between Design C and Design A is that the household would be asked to complete only the *Ongoing Expenditures Module*. Demographics were collected in the base survey. Income/assets, and larger and routine expenses, for the most part, are estimated from the *household profile component* of the overall sample. This component focuses on the smaller expenditures of a household.

Household Profile Component: Households in this component would be asked to keep records of aggregated expenditures at a modified 96-item level over two periods of three months. Most of this recording of expenditures would be proactive, using supported self-administration. At the initial interview, the field representative trains the respondent on the use of the tablet and its interfaces. The field representative asks the household to proactively keep receipts and use a supplied tablet computer to record expenditures during the upcoming quarter.

Household respondents would enter expenditure amounts in the tablet, indicating the expense category for the item. For example, a respondent might click on the category *men’s apparel and footwear* and then enter the amount spent for a pair of shorts or a pair of shoes. No further detail would be required. The level of entries is also reduced because expenses in some of the smaller and/or more frequently purchased categories (such as

food away from home, gasoline, nonprescription drugs, tobacco products, and personal care items) would not be collected in this component. Instead the household's expenditures for these items would be modeled from "like households" in the *detailed expenditure component*. Additionally a trip to the grocery store would require only saving the receipt or entering the total spent in the tablet. The allocation of food items purchased to details (such as meat, fruits/vegetables, nonfood items) would also be modeled from the *detailed expenditure component* for similar households. Thus the requested level of expense detail would be much less than in the supported journal component, requiring considerably less effort per week. Ongoing monitoring and feedback would encourage recordkeeping throughout the quarter. The field representative returns at the end of the quarter and conducts an interview (using the tablet) to fill information gaps. The goal is to end with a complete profile of expense data for the quarter at the 96-item expenditure data level, which includes some modeled components. The profile would include expenditures for smaller items, as well as breakouts of some aggregated amounts modeled from data from the *detailed expenditure component*.

The field representative would leave the tablet with the household for one additional quarter, reemphasizing the need to keep receipts and record expenditures on the tablet. The field representative returns again at the end of the second quarter and conducts an interview using the tablet to fill information gaps. At this point, he or she removes the tablet, ending the contact with this household.

The household respondent would also be asked to complete the *Income Module* on the tablet (described under Design A) at some point during the second wave to provide information on income and assets.

This component is similar to Design B and is modeled, in part, after the Westat proposal described in Chapter 4. It is similar to Design B in that it collects many of the data items in the *Recall Module* of Design B. It differs by placing the tablet with the respondent at the beginning of the reporting period and encouraging the respondent to report throughout the quarter. In Design B, the tablet is provided at the end of the period to the respondent, who is asked to use records and receipts to recall or estimate for aggregated expense items. Design C also differs from Design B in that reporting is for two three-month periods, as opposed to three six-month periods. Income, asset, and "labor force status" questions are included in the recall module. In wave 2, there is a series of questions about major life changes during the previous six-month period.

In both Design B and in the *household profile component* of Design C, a separate supported journal for detailed information is part of the design. In Design B, both pieces are collected from the same sample of households. In Design C, the details come from a different sample of households.

Frequency and Length of Measurement

All households would be interviewed in person on the base survey. Households selected for the *detailed expenditure component* would receive an additional in-person visit to place the tablet and initiate the supported journal keeping. These households would be asked to maintain a detailed supported journal for one month and to return the tablet by mail at the end of that month. Households in the *household profile component* would be interviewed in person two or three additional times. The first would be to set up the first quarterly recordkeeping period and place the tablet. Field representatives would make subsequent in-person visits at the end of the first quarter and again at the end of the second quarter. These last contacts might be made by telephone if the respondent is regularly entering expenditures throughout the quarter into the supplied tablet. Households in this component would be asked to report aggregated expenses over two consecutive one-quarter reporting periods.

Recall Period

In the base survey, respondents are asked to recall aggregated expenses for a variable recall period depending on the category of expense. In the two component samples, respondents are asked to record ongoing expenditures. Some recall questions regarding the previous quarter might be required to fill data gaps at the end of that quarter.

Incentives

Incentives would be used in both components of follow-on samples but not for the base survey. The panel envisions using the following incentives:

- \$150 per household for completing the *detailed expenditure component*, and
- \$180 per household for completing the six-month *household profile component*.

Role of the Field Representative

The field representative's role changes radically, from being the prime interviewer and data recorder to being the facilitator who encourages, trains, and monitors in support of the respondent's data entry. Ideally this will be true for both components. On the *household profile component*, the field representative may have to do more traditional interviewing if the household has not successfully kept up with recording expenditures over

the quarter. The role of the field representative would vary for households that use the paper-supported journal, and also across different tablet households depending on their comfort with the technology, willingness to use records, and household composition.

Role or Expectation of the Respondent

In Design C, the respondent takes on the primary role of providing the most accurate data possible, supported by records to the extent possible. What differs from the respondent's current role in the interview is that the respondent (after the initial meeting with the field representative) controls the time, pace, and order of data recording. In this sense, the design implicitly supports the idea that providing accurate data takes time and thoughtfulness. What differs from the respondent's current role is that the respondent, in the supported journal, is provided with augmented ongoing interactive encouragement and support from both the interface and from remote support staff.

Post-Data Collection Analysis

Estimates made using Design C would combine the strength of the larger sample size of the base survey and the more accurate detailed data from the follow-on components. They would be based on models that depend on the correlations between the estimates from the various data collections within each household and the correlations across aggregates of households based on household attributes. It is not expected that any one sample would stand completely on its own. The information collected would be strengthened through more detail or more breadth collected in other samples.

The panel envisions that the estimates of expenditures for the 96-item level aggregates and the more detailed estimates needed by the CPI would be model-based using data from both components and the base survey. A complete profile of individual households for microlevel research would be based on data collected for six months on the *household profile component*, which would be supplemented with estimates of smaller categories of expenses and/or the breakdown of collected aggregates using data from the *detailed expenditure component*.

Infrastructure

Infrastructure needs are similar to those in Design A and not repeated here.

Sample Size and Cost

The Design C sample sizes and costs (data collection only) presented in this section are broad estimates for use only to compare across alternatives. More careful calculations were beyond the information and resources available to the panel. These calculations are based in part on a spreadsheet of 2010 CE costs provided to the panel. The panel used those costs to estimate for similar activities, and then calculated an estimated cost per sample for the new prototype. Thus, these costs are for data collection only, and for data collection within a mature process. Estimates of sample size (net of 25% nonresponse) were then made that would keep to a neutral budget.

Assumptions:

- Cost per household in the base survey—\$324.
- Cost per tablet placement—\$165.
- 85% of households would use tablet. Costs go down with greater use of tablet.
- Cost of tablet—\$200. It can be used six times, with an expected loss of 10 percent.
- Remote monitoring of responses—\$100 per tablet household, for both components.
- In-person monitoring—
 - \$150 for the *detailed expenditure component*. Necessary for 10 percent of tablet households and twice per paper household.
 - \$200 for the *household profile component*.
- Incentives—\$150 per *detailed expenditure component*; \$180 per *household profile component*.
- Paper processing for paper households—\$100.

Based on these assumptions, the panel calculated the following projections for Design C:

- Total cost = \$20,600,000—above the budget neutral point.
- Annual effective sample size (assuming 75% response) for base survey = 25,000.
- Annual effective sample size (assuming 75% response) for *detailed expenditure component* = 11,000.
- Annual effective sample size (assuming 75% response) for *household profile component* = 7,000.

This prototype is shown with a cost projection that is greater than budget neutral cost of \$16,000,000. The base survey required considerable resources and the panel wanted to ensure that the quarterly panel

component had a sufficient sample size to be effectively used for economic research. The panel did not make projections on how high the correlations between the three components will be and estimated sample size without those assumptions. With relatively high correlations and modeling back to base, the precision of the estimates will increase. BLS should then be able to lower the sample sizes while maintaining the required precision, and thus bring costs back down. The extent of the increase in precision needs to be evaluated. The preliminary budget projections are

- Average cost per sample for base survey = \$324.
- Average cost per sample for *detailed expenditure component* = \$529.
- Average cost per sample for *household profile component* = \$962.

Meeting CE Requirements and Redesign Goals

This prototype meets the basic CE requirements laid out in *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011). Additionally, the redesign seeks to reduce burden, reduce underreporting of expenditures, and utilize a proactive data collection mode with newer technology. Table 6-3 provides more detail.

Targeted Research Needs

Most of the research requirements for this prototype are discussed in “Targeted Research Needed for Successful Implementation of Design Elements” on p. 178. Additional research is needed specifically for this prototype to:

- research and develop models for estimation using the base survey and two waves of data collection; and
- research and develop models for imputing at the household level “smaller expense items” collected on the *detailed expenditure component* and not on the *household profile component* into the household-level dataset to complete the overall household expense profile.

Comparison of Designs A, B, and C

As described above, the panel developed three prototypes of a redesigned CE. Each prototype is different—a good fit for a specific use of the data and perhaps less adaptable for other uses. Each prototype is similar in several ways. They all meet the basic CE requirements, take steps to reduce burden

TABLE 6-3 Meeting the CE Requirements and Redesign Goals with Design C

Goal	Design C—Dividing Tasks Among Multiple Integrated Samples
Produce quarterly and annual estimates of 96 categories of expense items	The design collects data at a more detailed level of expenditures than the 96 categories. Details are collected in the <i>supported journal component</i> . Larger and aggregated expenses are collected in the <i>household profile component</i> . Modeling data collected in both modules using correlations with base survey variables is expected to create estimates of detailed expenses with appropriate precision.
Income estimated over the same time period	Income is reported in the <i>household profile component</i> for each reporting period.
Complete picture for household spending	A complete picture of each household is collected for households sampled in the <i>household profile component</i> at the 96-category level. These estimates can be used independently or modeled along with data from the <i>base survey</i> and <i>detailed expenditure component</i> to enhance detail and precision.
Proactive data collection rather than change at the margins	The focus of this prototype is proactive self-reporting in both follow-on components. Expenditures are recalled in the <i>base survey</i> , but these data are not used for direct estimation.
Panel component with at least two contacts	There are two contacts in adjacent quarters for each household in the <i>household profile component</i> .
Reduced burden	This prototype reduces burden by dividing the overall tasks among multiple integrated samples, and not asking each household to perform each task. Some households are asked to provide detail over a relatively short period. Other households are asked to report over a longer period, but are asked for less detail. Incentives are available to reduced perceived burden. Burden is reduced by making the response tasks easier and more intuitive with the tablet interface and support. Incentives are used to reduce perceived burden.
Reduced underreporting	The <i>detailed expenditure component</i> focuses on detail, the increased use of records, allowing the respondent to record expenses at a time best suited and in a way best suited to him/her, reduced proxy responding, and incentives to perform the task. All are expected to lead to a reduction in underreporting detailed expenditures. The <i>household profile component</i> also focuses on use of records, asks for less detail, and allows the respondent to record expenses at a time best suited for him/her. The ability to model using data from all components will allow for better adjustment of underreporting when it is discovered.
Budget neutral	The sample sizes that are calculated exceed the budget neutral level. If the correlations between the base variables and the two components are sufficiently strong, these sample sizes could be reduced.

and to incorporate technology, and use as part of their design an implementation of self-administration with support from the field representative, a tablet computer, and a centralized support facility.

To recap the main features of the three designs:

- *Design A—Detailed Expenditures Through Self-Administration* is designed to provide all of the detail that is needed currently for the CPI. It maximizes the use of self-administration through a supported journal for concurrent data reporting. It also features a self-administered recall component to collect larger and recurring expenses. It stresses that the recall period should be the one best suited for accurate reporting, which might vary by expense category. Cautions: Diary fatigue is an issue with the current Diary survey. The supported journal is designed to address this issue, but research is needed to develop and investigate this hypothesis. Additionally, Design A is not as adaptable for economic research as are the other two designs. It provides panel data with two waves of collection from the same household three months apart. The use of varying reference periods for different expense categories is likely to make the formation of a consistent household dataset for research more difficult.
- *Design B—A Comprehensive Picture of Expenditures and Income* focuses on providing a rich dataset that will meet the needs of economic research. It uses a recall survey to collect expenditure information over the previous six months. Incorporating three waves of data collection with the same household, it results in a panel dataset covering 18 months for each household. Waves 2 and 3 would be self-administered. The questionnaire would ask about aggregated (96-item level) expenditures, rather than at the more detailed level of the current Interview survey. It combines two collection methods: (1) asking respondents to recall specific expenditures for larger and recurring expenses and (2) asking respondents to estimate the “average or typical” amount spent on other types of expense items. This design also incorporates an intensive subsample in which households will be asked to work with the field representative to balance household income and expenses over two one-year periods. Cautions: A three-month recall period caused issues with the current Interview survey, and this design uses a recall period that is twice as long. Research is needed to see if the new approach (aggregated expense items, longer recall period, and increased use of respondent estimation of “typical or average” expenditures) will provide an acceptable level of accuracy. Design

B does not provide expenditure estimates at the level of detail currently used by the CPI.

- *Design C—Dividing Tasks Among Multiple Integrated Samples* provides both the detailed expenditures required by the CPI plus a consistent and complete picture of household expenditures for economic research. It uses a base survey for stratification and sampling, followed by subsamples for a *detailed expenditure component* and a *household profile component*. Estimates of detailed expenditures would be made using data from all three components through modeling. The *household profile component* provides a consistent panel dataset for six months and is designed for use for economic research. A richer dataset for research could be developed through modeling with the other components. Cautions: Because of the base component, Design C may be more costly than the other two designs. If the correlation structure between the components is strong, sample sizes could be reduced and this might not be a problem. This design is complex and will take more resources to develop and test the required models. The panel data are provided for only six months rather than a full year.

The prototypes discussed above use a variety of methods for collecting expenditure data. Variations in the exact form of the collection methods and the relative emphasis across methods are what distinguish the prototypes from each other. The major collection modes are¹

- *a module to collect demographic and socioeconomic data, as well as information on life events.* These modules reflect, inter alia, the need to classify households for all of the major purposes of the CE.
- *a recall module.* Design A includes a recall component to collect information on large and routine expenses. Designs B and C include recall components to collect spending for 96 expenditure categories. The recall periods are unspecified and variable for Design A and are used as needed to fill gaps in concurrent reporting in Design C. Design B has a six-month recall period, but it allows for a shorter period for some spending categories where spending is likely to be relatively stable over time. Design B explicitly accepts as (almost) inevitable that respondents will estimate spending patterns over the recall period.
- *a supported journal.* Designs A and B include supported journals

¹The intensive panel discussed under Design B is not discussed in this section, because none of the options requires that this approach be successful. Moreover, all of the options would benefit if it is found to be effective.

to collect all expenditures over two- and one-week periods, respectively. The comparable supported journal for Design C would cover one month. Design C also includes a six-month supported journal to collect spending on the 96 expenditure categories covered by the recall modules in Designs A and B.

As emphasized above, the panel has no empirical evidence on how well these collection modes—and the variations within them across prototypes—will work. In the current CE, both the Diary and Interview surveys have serious shortcomings. Comprehensive and sophisticated—and, unfortunately, expensive—testing will be necessary to determine which, if any, of the collection modes in the prototypes can be successful. Among the key questions to be answered with respect to recall and the supported journal in the collection of expenditure data are the following.

Recall components:

- Can a recall survey be designed that could collect sufficiently accurate data on spending patterns for individual households at the level of 96 expenditure categories?
- How does the length of the recall period affect the efficacy of a recall survey?
- Is it necessary to accurately reconstruct actual spending or would it suffice for a household to estimate its spending patterns?
- How can new technology, including self-reporting on a tablet, be best used to improve a recall survey?

Supported journal:

- Can a supported journal overcome the diary fatigue and under-reporting that has plagued the current diary instrument?
- For how long can households be asked to complete a supported journal?
- How would a supported journal that asks for spending aggregates, rather than individual purchases, work in practice?
- Can a structured journal be designed that will reduce respondent burden and increase and sustain the level of reporting?
- How critical a role would tablet reporting play in the design?

The answers to these (and other) questions will be critical to the CE reform effort and how the data from a new CE survey can meet the requirements specified by BLS. The implementation risks are significant:

- If it proves unworkable to collect expenditure data by recall, the

spending data collected from each household—the number of expenditure weeks—will be seriously curtailed, increasing the variance of CPI weights and publication aggregates and making it impossible to construct annual spending patterns at the individual household level.

- If a supported journal cannot sufficiently mitigate the problems with the current diary, it will be impossible to get spending estimates at a fine level of category detail.
- The six-month supported journal in Design C is the most innovative of the collection methods in the three prototypes. It could be an alternative to a recall survey, albeit with compromises on the length of time covered. However, it requires modeling daily spending to 96 categories of expenses for six months; there is no experience on which to gauge how well an operational version can be designed and implemented.

The panel made an effort to compare data collection costs across prototypes. This proved very difficult to do, and the panel had to make many assumptions and guesses. Thus these cost estimates must be considered as very preliminary, and used only as a general comparison across prototypes. More precise costs would require considerably more information and resources than the panel had available for its deliberations. However, they are useful in comparing the three prototypes and in calculating the very approximate sample sizes that might be possible in the current budget environment.

BLS provided the panel with basic costs on data collection from the 2010 CE. It was difficult to extract the costs for initial screening for out-

BOX 6-1

Example: Calculation of Average Cost per Sample for Design A

Cost per household using a tablet:

Tablet placement (twice) + Prorated cost of tablet + Remote monitoring + In-person monitoring + Incentive

$$\$330 + ((\$200 + \$20) / 6) + \$200 + (\$150 * 0.1) + \$200 = \$782$$

Cost per household not using a tablet:

Paper diary placement and pickup (twice) + In-person monitoring + Paper processing + Incentive

$$\$660 + \$300 + \$100 + \$200 = \$1,260$$

Average cost per household, assuming 85% use tablet:

$$(\$782 * .85) + (\$1260 * .15) = \$854$$

of-scope households and for contact with households that refused. Thus the panel focused on costs for completed interviews/diaries, making the assumption that the cost for screening and refusals would not vary greatly between prototypes. Following several conference calls with BLS and Census staff, the panel approximated the “budget neutral” amount for data collection for completed interviews/diaries as \$16 million. This figure may not be very accurate, but it gives a base from which to make projections.

The panel used component costs provided by BLS for completed interviews and diary placements as a guide to provide very basic cost parameters (assumptions) for each prototype. The panel speculated on other costs. These cost parameters are provided in the report with the description of each prototype under *Sample Size and Cost*. The cost parameters were combined to create an estimated “average cost per sample” for each prototype. Box 6-1 provides the example of calculating this cost for Design A. This average cost was divided into the total cost to calculate an annual effective (completed) sample size. This number was rounded to avoid suggesting greater accuracy than this process could produce.

Table 6-4 provides a comparison of cost and possible sample size between the current CE and the three prototype designs. It is extremely important to note that these costs are for data collection only and for a mature data collection process. Any major redesign has many additional

TABLE 6-4 Comparison of Key Elements of Prototypes with Current CE

	Design A	Design B by Components		
		Base	Intensive Subsample	Total
Sample size—Households (respondents)	18,700	12,200	800	12,400
Total collections across waves (respondents)	37,400	36,600	1,600	38,200
Weeks covered by direct reporting of expenditures	448,800	878,400	0	878,400
Incentives per household	\$200	\$300	\$300	\$319
Cost—Total (respondents only—75% response rate)	\$16,000,000	\$13,900,000	\$2,100,000	\$16,000,000
Cost—Per responding household	\$853	\$1,138	\$2,625	\$1,311

SOURCE: Developed by panel.

costs that are not included here, including costs for development, research, testing, and retooling.

The sample sizes presented here are very general estimates of the effective sample sizes (net of a 25% nonresponse) that might be possible while remaining budget neutral.

To summarize, it is important to emphasize that extensive testing is necessary to determine which of the collection methods and their variants contained in the prototypes can be made operational. The prototypes embody the panel's informed opinion about the best avenues to explore, not options from which to choose a survey design.

Using the Tablet PC for Self-Administered Data Collection

The panel proposes the use of tablet computers with wireless phone cards for supported self-administered data collection in all three prototypes as a way to solve some key problems with the CE. Lightweight and easy-to-use tablets represent stable (robust) technology, are commonplace, feature more than sufficient computing power, and are economical in price. In fact, their price continues to fall over time. This section provides more detail on the use and management of these tablets and the support system for their use.

For this data collection effort, a low-end tablet is likely to suffice. Their

Design C by Components				Current CE by Components		
Base	Supported Journal	Household Profile	Total	Diary	Interview	Total
25,000	11,000	7,000	25,000	7,100	7,100	14,200
25,000	11,000	21,000	57,000	14,200	35,500	49,700
0	44,000	168,000	212,000	14,200	340,800	355,000
0	\$150	\$180	\$108	0	0	0
\$8,100,000	\$5,800,000	\$6,700,000	\$20,600,000	\$2,833,822	\$13,086,524	\$15,920,345
\$324	\$529	\$962	\$824	\$399	\$1,843	\$1,121

use by respondents could reduce the number of field representative visits and associated field costs. The resultant savings could be used to fund worthwhile incentives to facilitate respondent participation. Through self-administration, interviewer effects would be minimized and data would be recorded on an ongoing basis, substantially reducing recall bias. Reliance on proxy reports could be reduced, as well. Since many people successfully use advanced self-administered applications on the Internet, adoption of a self-administered tablet data collection mode would not require large respondent training costs, especially if the data collection software were designed to be user-friendly.

These prototype designs would require an initial in-person visit from the field representative. The field representative would explain the survey, secure cooperation, deliver the tablet and demonstrate its use (i.e., train the respondent), explain the incentive structures, and answer questions. For some prototype designs, the field representative assists the respondent in self-administration of the initial survey module as a way of both collecting information and training on the tablet. The field representative would leave the tablet computer with the respondent, as well as a package in which to mail back the computer after data collection is completed. Self-administered data collection subsequent to the initial visit for the household would utilize the tablet to the greatest extent possible.

Technology Assumptions

This approach adopts established, relatively current technology. The only assumptions made about future technology are that (1) the wireless phone infrastructure will continue to expand and become faster, and (2) tablet computers are a viable platform for the foreseeable future (i.e., the next 10 years). Development software must be adaptable to future platform changes, but that is not expected to pose a problem.

The tablet can be situated in the respondent's kitchen or on the dining room table, accessible to any of the household members charged with entering data. To be user-friendly, it will need to have a fast boot-up protocol, and all CE applications will have to execute very quickly with self-evident interfaces. Respondent training is expected to be short and to focus on the task completion (i.e., content and entry of data), rather than navigating the technology.

The tablet would house only necessary applications including logon, encryption, connection to the server, and the electronic instrument itself. As data are entered, they are sent to the server in real time. Such a system requires that the tablet be "locked down." That is, it cannot be used for any other applications and non-CES applications cannot be loaded onto it. The sole exception would be if a limited number of popular free game

applications were installed as “incentives” to respondents to familiarize themselves with and use the tablet regularly.

The data collection application must have a simple, high-speed interface. For this reason, the application software is held on the tablet. Data are transferred immediately to the server. Data do not remain in storage on the tablet between activities with the respondent. Only secure data are communicated to the server, and not interface pages, routing information, or other application rules. A respondent can review the history of household entries at any time, and those data would be transmitted back to the tablet for review.

Respondents must be certain that their data are secure. Hardware, software, and procedures are built around secure protocols. Both the field representative and the technical support desk must be able to clearly explain how these data are held securely, in one place, on the data collection server. Respondents must also be aware that the use of these tablets represents better, more secure data collection than alternatives while reducing data collection costs.

Because relatively little consumer data are sent to the server at any point in time, wireless phone infrastructure is being proposed for communication. The coverage in the United States is extensive and ever increasing. Coverage for the CE target population is likely to be virtually 100 percent within several years. The use of this communication infrastructure means that a household’s Internet connectivity is irrelevant. Moreover, the link-up to the server would be immediate and transparent to the respondent.

Monitoring and Technical Support

Since data are received upon entry for tablet users, the respondent’s cooperation and compliance can be tracked by a centralized facility. For those who respond daily, it is possible to send positive reinforcement (e.g., a thank-you message) every time they log in. For those who lapse in their task, it is possible to intervene in a timely and effective manner with supporting reminders via phone, e-mail or text, or by mail. Ideally, a 24/7 in-house technical support team would be maintained throughout the data collection period to address respondent messages and phone calls very quickly.

The field representative remains involved with the household as needed. The representative may revisit some households to assist with or to perform the data collection. The representative may need to retrieve the tablet from some households after administration concludes, although this would be minimized by the distribution of the return-mail packages. However, these actions would take place only if requested by the central facility.

The tablet would be lightweight, have a large and bright enough screen

to be readable by even those with vision problems, and would be sturdy. It would withstand many mailings between the respondent and the home office. At every trip to the home office, the tablet would be inspected, cleaned, and recertified for continued data collection. Tablet software could be updated when it is in the home office.

Implementation Program

A rich implementation program will inform the many details associated with this kind of data collection. Any research taken as part of this program needs to be focused on optimal approaches, not on technical feasibility. Research will be needed on human interactions, including interface, incentives, training, protocols, and support. Development of the necessary in-house production infrastructure would include server development, communications, and sample management/tracking systems.

The implementation program would be iterative. Goals for key rates (e.g., rate of acceptance, rate of cooperation, and tablet longevity) would be set and monitored. Shortfalls would be quickly addressed via tailoring, using modified or new approaches. More details are provided in “Targeted Research Needed for Successful Implementation of Design Elements” on p. 178.

The Respondent

Under a successful implementation, most respondents would be able to execute the tasks via tablet within a few minutes. A well-designed interface would make long training sessions unnecessary. Field representatives would be able to assess respondents quickly to determine the most appropriate training or data collection approach.

Respondents are busy and do not have time to deal with slow interfaces and software awkwardness. The model for the interface would be similar to that of TurboTax rather than the linear flow methodologies of today’s self-administered questionnaires. This means that the app would be built in modules, and the user could choose to fill in the information in any order that is convenient. The user can also come back to any item and make a change at any time. Alternatively, the user (respondent) could choose to use a structured interview approach for moving step-by-step through the app.

The use of the tablet would be as simple as the following:

1. Touch the on button.
2. Log on.
3. Observe the welcome and messages (can get past this easily).

4. View the data collection module, which pops up fast, is easy to use, and is self-evident (visually coherent).
5. Enter data, with computer response time being virtually instantaneous.
6. When entry has been completed, hit the off button.

Delivery and Mail

The first time the tablet enters a household, it is brought by a field representative. The field representative leaves behind the tablet and a mailer package, and explains the incentives. After the recording period, the respondent slips the tablet into the mailer and sends it back. In subsequent waves, the tablet is mailed to the household and back to the home office. Since no data are held on the tablet, there is no security issue.

Under each of the three prototype designs, the panel estimates a loss rate of 10 percent per year for tablets. (This loss figure is stated as part of the sample size and cost assumptions under each of the three prototypes.) This estimate is relatively low because a tablet configured to allow only authorized survey activities would be of little later use to a household. However, in the next subsection, “Incentives and Cooperation,” the panel presents an option to allow some free game applications on the tablet as a way to encourage familiarity with the tablet. If this is done, then the tablet would become more valuable to the household and the loss rate could be higher. BLS should consider these issues and may want to hold final incentive payments to households until the tablet is returned. As an alternative, BLS could consider the tablet to be part of the incentive package and plan to leave it with the household. It is clear that the alternative of sending a field representative to pick up each tablet would add significantly to the survey cost.

Incentives and Cooperation

Incentives and monitoring are integrated. Respondent behavior is monitored constantly. Daily logging on is encouraged through the use of participation incentives and other forms of positive feedback. Longer-term engagement could also be enhanced through use of popular free game applications on the tablet. If the respondent fails to log on for a period of time, then an immediate automated intervention would be issued. This would start with the field representative making a call and offering assistance. If necessary, the field representative could revisit the household to encourage the use of the tablet, or to switch the respondent to an alternative mode of collection. The field representative would have access to respondent tracking reports so that intervention could be timely and effective.

Home Office Support

From the perspective of the survey organization, these prototype designs shift activities to a more centralized basis, allowing more consistent, frequent, and responsive monitoring of data collection from a centralized facility. The role of the field representative is recast to one of gaining cooperation, training, modeling use of the tablet, and providing field-based support, rather than being solely an interviewer or supported journal-purveyor in a traditional sense. This change provides savings on field costs. Technical support staff would be available to respondents 24/7 (ideally) and would typically answer a call within a few rings. Possibly the support person can see the same screen as the respondent sees.

Tracking and assessment algorithms would run on the server in order to detect patterns of deception, respondents not cooperating, and so forth. Any issues would result in an intervention. Cooperation would be also quickly recognized with positive reinforcement. Incentives would be handled quickly, perhaps on an ongoing basis.

Incorporating Cognitive Changes to the Paper-Supported Journal

The panel recommends that a paper instrument be available for use by respondents who cannot or will not use a tablet computer to record and submit expenses. The current Diary booklet should not be used and the panel recommends a redesign (Recommendation 6-5). Considerable improvement can be made in the current Diary to ease comprehension and use. In addition, the structure of the tablet-supported journal and new paper-supported journal need to be aligned to the extent practical so as to minimize mode effects in data collection.

Recommendation 6-5: A tablet computer should be utilized as a tool in supported self-administration. However, a paper option should continue to be available for respondents who cannot or will not use a tablet computer. Visual design principles should be applied to redesigning the paper instrument in a way that improves the ease of self-administration and is aligned with the tablet modules.

With an eye to redesigning the paper version, the panel concludes that the current diary lacks a clear navigational path and linear flow that allow respondents to proceed page by page through the diary completion process. Currently the instructions are located throughout the booklet, where they seem to fit the page structure but force the respondent to search for information. The panel recommends eliminating “daily” report pages. The breaking apart of each category by day of the week and overflow pages

creates a larger diary booklet and makes it much harder for respondents to review the diary and see whether they have forgotten to add anything. With current content, it may be more effective to color code different sections of the booklet for each category of expenditure, and allow household members to keep a continuous listing of each expenditure category for the week. This structure would also work better for individuals who collect receipts and do not record expenditures each and every day, an occurrence that seems increasingly likely as alternative ways of purchasing and paying for items continues to grow.

The panel recognizes that the Diary has been revised in recent years, and the current format is considered a substantial improvement over previous versions. However, visual layout and design principles have continued to improve, and they need to be applied to producing a new paper-supported journal. This will require research aimed at testing to produce as much consistency as possible between how people respond to the tablet and paper forms of data collection.

ROADMAP TO GET THERE

The road to completing a redesign of the CE is difficult but has a very important destination. BLS began this journey in 2009 with the initiation of the Gemini Project. This report from the National Research Council provides some additional mapping for the future. The panel sees a number of steps ahead:

- Prioritize the uses of the CE.
- Conduct a feasibility study of relevant redesign protocols.
- Make redesign decisions based on the prioritization and assessment.
- Incorporate the use of newer technology as appropriate in the new design.
- Update the capabilities of BLS staff and outsource to find needed expertise.
- Conduct research that is targeted to successful final implementation of design elements.

This section provides some additional guidelines for these steps.

Timetable and Priorities

It is not the panel's role to impose a specific timetable on BLS for investigation or implementation of the alternative designs discussed here. However, the panel believes that development of a targeted and tightly focused plan is necessary if BLS is to achieve a redesign within the next five years.

To that end, the panel suggests a prioritization of research and development activities and recommends that BLS adopt an ambitious timetable for working toward an alternative design.

A key initial step is to identify and prioritize the uses of the CE. A key finding of the panel is that many of the problems faced by the CE can be attributed to the multiple competing demands on the survey. In trying to be all things to all people, and to achieve maximum breadth and depth, the CE surveys have become unwieldy and increasingly unreliable. Without a prioritization of key purposes of the CE, and a corresponding acknowledgment that all purposes will not be equally well met, any redesign is unlikely to be a successful venture. Making the difficult decisions about what is to be sacrificed in order to improve the CE overall is a necessary first step on the redesign path. Without this, the remaining steps are unlikely to yield a successful outcome. The panel recommends that BLS undertake this process of the prioritization of key uses, with buy-in from stakeholders, as soon as possible, and believes that it can be done within six months.

Another key element of the panel's recommendations is that some form of tablet-based instrument, used for both recordkeeping and reporting of expenditures, as well as for more traditional computerized question-and-answer process, is an essential ingredient in a new design. A number of key untested assumptions need to be addressed before proceeding with using this tool in alternative designs, including (1) whether the use of tablets will reduce burden and improve the quality of reporting over the paper-based diary and the interviewer-administered quarterly recall surveys; (2) what proportion of households will be willing to use tablets; and (3) if the introduction of tablets will reduce resistance to participation on the CE such that the overall field effort is reduced without negative effect on response rates. Answering these questions is critical in order to proceed with any of the alternative designs. A first priority is to design, build, and test a prototype tablet instrument. The panel believes this can be done within two to three years, but is likely to require outside help, especially from those with experience building tablet applications, and those familiar with the hardware and software issues related to deploying this technology.

If the tablet-based data collection approach proves feasible, other research and operational questions remain. They include questions such as whether varying the recall length for different expenditures would be effective to reduce burden and improve data quality; whether asking about broader categories of expenditures rather than the detailed items would similarly improve reporting; and how to structure incentives to maximize response. Some of these are described in more detail in the sections below. The panel views these as important, but contingent upon the successful implementation of a tablet-based approach.

Again, the panel is of the opinion that a precise timetable from them

(the panel) for implementation would not be helpful to BLS at this stage. Rather, it is recommended that BLS lay out a path to implement one of the three alternatives (or a different combination of the components of these three designs), with key decision points along the way. The feasibility studies outlined in Recommendation 6-3 would be a key component of this plan, and would be a key assessment on which to base further decisions.

These and other decision points need to be identified along with a clear understanding of the potential outcomes from research that will drive those critical decisions. Additionally, the alternative actions stemming from the decisions need to be clearly articulated. In other words, BLS should identify criteria for success (or continued research and development) at each critical juncture in the process. This will determine whether sufficient evidence exists to proceed to the next stage of development, or if alternative paths need to be followed. Some of the work can be done in parallel with work on the critical path, but it should not detract from the resources focused on addressing the key issues and reaching the key decision points within a reasonable time frame.

Recommendation 6-6: BLS should develop a preliminary roadmap for redesign of the CE within six months. This preliminary roadmap would include a prioritization of the uses of the CE, an articulation of the basic CE design alternative that is envisioned with the redesign, and a listing of decision points and highest priority research efforts that would inform those decisions.

Guidelines for the Use of Incentives

The purpose of incentives is to provide motivation to the respondent to complete a data collection activity and to take the time to provide accurate data. The CE in its current state and its future redesign will require considerable effort from respondents. It is the panel's opinion that an appropriate incentive program will be needed as a part of this program. Monetary incentives are used in federal surveys but their use is not common. The Office of Management and Budget must approve each use of monetary incentives and look for strong justification based on criteria such as improved data quality, reduced burden, and/or improved coverage of specialized populations, and/or for surveys that are particularly complex (U.S. Office of Management and Budget, 2006). The panel believes that these criteria can be met by a carefully designed incentive structure for the CE.

This section provides a short background on incentive use in surveys and some guidelines for developing an incentive structure for the CE. The exact details of that structure, the exact kind and amount of incentive, and

the appropriate placement of incentives within the data collection window can only be determined with appropriate CE-specific research.

Overview and Basic Guidelines from Survey Research

The use of “incentives” is a standard and accepted component of many survey efforts in the United States. As Singer (2002, p. 3) appropriately described, “Incentives are an inducement offered by the survey designers to compensate for the absence of factors that might otherwise stimulate cooperation.” Interesting, however, there is less agreement within the research community regarding why incentives work, or at least no single theory describes when and why some incentives work and others do not. Some view an incentive as a “social exchange” between the researcher and the respondent. By providing something of value to the respondent, the respondent should, in turn, provide his or her cooperation. Others view it as an economic exchange, whereby the respondent views the offer to participate in terms of “costs” and “benefits,” and use of incentives can help to improve the perception of “benefits.” In reality, both factors are likely at work, with differential effects of each seen across different studies, populations, and contexts.

Typically, researchers use incentives to achieve one or more of the following goals: (1) improving overall response rates, (2) enhancing the characteristics of an unweighted set of survey respondents, (3) decreasing the likelihood of missing data or other factors that affect data quality, or (4) reducing the total costs of fielding a survey (Brehm, 1994; Church, 1993; Dillman, Smyth, and Christian, 2009; Singer et al., 1999).

In practice, incentives are sometimes used uniformly across all sampled units/respondents, or alternatively they may be used differentially across populations, contexts, modes, etc. In the first instance, all households/respondents receive exactly the same form, level, and timing of incentive. This is often done either for reasons of “fairness” (trying to treat all respondents identically) or for operational efficiency (simpler to execute and track). The downside is that incentives may be used where they are not really needed and/or the amount required to obtain participation from a particular subgroup or context may be insufficient. The use of differential incentives, in contrast, is based on the premise that incentives should be targeted to populations or points in the survey process where burden is highest or the likelihood of response is lowest (i.e., where task or burden may result in differential nonresponse) (Link and Burks, 2012; Martinez-Ebers, 1997). Use of differential incentives is often justified, therefore, based on effectiveness, efficiency, and “need” basis, but can be criticized for no longer treating all sampled units identically.

In terms of form and timing, incentives come in a variety of forms and

are most effective when targeted at the points in the process where they will be most effective. Numerous studies have examined the efficacy of different types of incentives, such as cash, checks, cash cards, sweepstakes, points and gifts, virtual rewards (e-badges, access to unique online content or functions), donations to charity, e-coupons, and so on (Antin and Churchill, 2011; Balakrishnan et al., 1992; Bristol et al., 2011; Trussell and Lavrakas, 2004; Warriner et al., 1996; Zichermann and Cunningham, 2011).

While cost is always an important factor, considerable care needs to be taken in matching the appropriate form of incentive to the specific population of interest to achieve the desired goal or outcome (i.e., cooperation, long-term compliance, higher data quality, etc.). There are also temporal aspects to the use of incentives. They may be paid upfront at the time of the survey request (“non-contingent incentives”) or paid upon completion of the task (“contingent incentives”) (Bensky et al., 2010). Incentives may be used at different points in the survey process—for recruitment, at the start of a survey, after the survey, or partial over time in the case of a panel or longitudinal effort. The value or form of incentive can also be different at these different junctures. Again, the researcher needs to consider carefully the form, amount, and timing of the incentives throughout the data collection process to achieve the desired study goals.

Cost is an important consideration in the use of incentives. Some researchers view incentives as an “additional cost” to their study design and will exclude them or cut them at the first sign of budget issues. In reality, if used effectively, incentives can help to reduce costs and become an essential component of the overall survey design (Brennan, Hoek, and Astridge, 1991). If used effectively, a modest incentive can often gain cooperation from a respondent far less expensively than having an interviewer try to convince a respondent to participate. Use of an effective incentive design can, therefore, reduce more costly interviewer time and/or achieve a higher rate of participation than when incentives are not used.

Guidance for the Consumer Expenditure Program

It is critical that some form of incentive structure be put in place in the redesigned CE regardless of what develops as the final data collection design. The reasons for this are cost-efficiency (e.g., using incentives to help reduce field time, labor costs, and other expenses) and as an offset of respondent burden. The exact form, amount, placement, and timing of the incentive structure for any given design will require pre-deployment research and testing. While lessons can be learned from a thorough review of prior published studies on incentives, the one truism in the use of incentives is that there is no single “magic bullet.” The effectiveness of a given incentive design is directly tied to the population of interest, nature of the

data collection request, modes of data collection, length of participation requested, and other factors. Prior studies do, however, provide a guide in terms of a starting point and most effective approaches:

1. Cash is always the most effective motivator—monetary incentives, particularly when paid in terms of cash (as opposed to check, cash card, or e-payment), have a more positive effect than nonmonetary incentives even when the relative value is equal.
2. Prepaid incentives work better than promised or post-paid incentives—an incentive provided to the respondent at the outset of the survey request appears to have a much greater impact on participation than does the promise of an incentive upon completion of a task, even if the value of the up-front incentive is somewhat smaller than the amount of the promised incentive.
3. Incentives are most (and sometimes only) effective if utilized at the correct points in the survey process. For instance, if an incentive is offered for participation, it will be most effective if mentioned at the outset of the recruitment conversation as opposed to after the respondent has effectively agreed to participate.
4. In panels or longitudinal efforts, incentives should be used throughout the process to maintain compliance, rather than provided entirely up front or at the completion of the panel. Ongoing rewards, both tangible and intrinsic, are important for maintaining long-term participation.
5. Use of differential incentives (e.g., providing differing incentive forms, amounts, or timings to different sets of individuals based on varying degrees of burden or likelihood of cooperation) should be seriously considered for any incentive structure to optimize the efficient use of these funds.
6. Larger incentives (\$100+) should be considered in any instance where the data collection request is particularly intrusive or sustained.
7. Intrinsic incentives (i.e., nonmonetary motivational approaches) should be a part of any panel or long-term data collection effort to help sustain respondent interest and motivation. These are a range of incentives of this type, including use of “e-badges” (electronic notices in the form of a badge) for completion of successful tasks or meeting milestones; status levels (i.e., silver, gold, platinum) for reaching critical milestones or longevity; and random notices congratulating or thanking the respondent for their participation (Antin and Churchill, 2011; Zichermann and Cunningham, 2011). These types of incentives are easily developed within a tablet envi-

ronment; however, there is little research currently on the effectiveness of these types of strategies on data collection activities.

Incentives for the Three Prototype Designs

Each of the three prototype designs discussed in this chapter recommends the use of incentives. The type, amount, and placement of those incentives need to be researched and tested. The details proposed under each prototype give an overall sense of the likely size of the incentive and the need to incorporate incentives into the overall cost structures.

In developing the prototypes, three additional ideas surfaced and are included here to generate further discussion about potential motivation of respondents in the CE.

- Provide households with a financial profile, comparing their expenditures with other households in the same income bracket and demographics;
- Install a limited number of popular applications (for functional use or to play games) on the tablet as an “incentive” to respondents to familiarize themselves with and use the tablet more regularly; and
- Give the tablet to the household as the post-paid incentive at the end of the more intensive data collection panels.

Recommendation 6-7: A critical element of any CE redesign should be the use of incentives. The incentive structure should be developed, and tested, based on careful consideration of the form, value, and frequency of incentives. Serious consideration should be given to the use of differential incentives based on different levels of burden and/or differential response propensities.

Guidelines for Adopting Newer Technology and Incorporating External Data

The panel has proposed three prototype designs for the CE that make use of a tablet computer. Those prototypes do not include recommendations for use of any specific external datasets. However, the panel encourages BLS to explore other technology and administrative data sources as they move into the future. This section provides general guidelines on this topic.

Adopting Newer Technology

It is important for the CE to pursue a research agenda that explores and adopts new technology and considers the utility of public and private administrative records. However, such an agenda requires discipline since neither technology nor administrative records ought to be pursued for their own sakes. It is important for the agenda to promote a *strategic direction* for continuous improvement, creating reductions in:

- *Data collections and processing costs*—respondents entering supported journal data via computer could reduce data collection and processing costs; similarly, the use of administrative data could reduce the amount of information that is needed from the respondent, further reducing data collection costs;
- *Measurement error*—reductions can be achieved by tailoring the technology to the user, whether the user is the respondent (completing a supported journal) or the interviewer (administering a questionnaire); coupling this with the collection of aggregated data (i.e., collecting less detail about purchases) will ease burden for both the interviewer and the respondent; and
- *Statistical variance and complexity of the CPI estimate*—the incorporation of administrative data in the CPI estimation process could in principle result in increased statistical precision and reduce data collection costs by requiring that fewer items be collected.

Two additional factors are important to be incorporated into an agenda that considers technology and administrative records for the CE. The first is robustness of technology. Releases of new technology (both hardware/firmware and software) have accelerated over the last decade, and this trend is expected to continue. Relatively old technology such as palm-pilot computers have been eclipsed by smart phones, and the speed and capabilities of smart phones are growing rapidly from year to year. And while laptops have found a place in society for over two decades, tablet-style computers (which are also relatively old technology but only about half as old as laptops) have begun enjoying a recent upsurge in development and use. These are but two examples of how quickly the landscape of technology is changing. The choices of equipment for CE data collection must be sensitive to such change. Because the pace of technological development is expected to continue to accelerate for the foreseeable future, the CE would not be served well by selecting a technology that would be unavailable or no longer supported by the time testing and piloting have been conducted. Instead, the CE needs to identify and test technology that is robust with respect to technological advancement.

The second factor lies in risks associated with the indeterminate availability of administrative data. While administrative data hold promise of creating great efficiency, their perpetual availability over time cannot be guaranteed. The biggest risk would be to identify a solution to the CPI that relies heavily on administrative (auxiliary) data, only to have that source of data become unavailable. There are real risks of this in the current volatile economic environment. Governmental budget cuts and/or private sector changes, such as business closings and release restrictions of sizeable magnitudes, are to be expected in the coming years. And societal change, including heightened concerns over privacy, could limit the availability of data that are currently accessible from other sources. There is also concern about potential liability risks for BLS to access and store individual financial data.

It is important for BLS to pursue the exploitation of administrative records. However, the panel believes that it would be problematic for BLS to convert irreversibly to full reliance on such solutions unless there is full confidence in the perpetual availability of those data. In the absence of such confidence, the prudent approach would be to maintain some level of primary data collection as a hedge against the risk of the loss of administrative records, as well as a checking mechanism to validate the quality of the administrative data.

The integration of technology and/or administrative records into the CE (and the CPI) is best accomplished as a natural part of continuous process improvement. A transition window of five to seven years could be utilized for new technology absorption into CE design and operations. Potentially advantageous technologies could be identified and prioritized on an annual basis, followed by a process of feasibility testing, assessment, planning, piloting, and adopting into CE operations. Such a process requires that the CE design maintain experimental panels integrated into the sample design specifically for the purpose of field-testing and evaluation.

Incorporating External Data

The panel developed recommendations regarding the use of extant administrative data for the CE. The potential use of external records or alternative data sources as a replacement or adjunct to current survey data for the CE is often raised in discussions of a CE redesign. Whether at the aggregate or the micro (respondent/household) level, the appeal of “readily available” information is that it appears, at first glance, simple to use. Although such information might hold great promise, we also realize that such use is accompanied by corresponding great risk, particularly from a cost/quality trade-off perspective. That said, there are scenarios under which these data could be quite useful, in particular as category-specific

quality assessment tools at the aggregate level and “memory joggers” at the household level.

Use of aggregate data. Aggregate retail data from scanner receipts can provide greater chronological and item detail, with potentially greater accuracy, than traditional data collection methods—offering opportunities for significant quality improvements for CPI budget shares and aggregate spending tables. The potential benefits include more data with less variance, as well as better data, leading to less bias. That said, use of aggregate data is not likely to yield sufficient coverage of all goods required by BLS; therefore, aggregate data cannot replace the current interview process in whole. Several areas need further exploration, including use of aggregate data in the CE process to (1) replace some of the detail (and hence burden) of the current CE interview process, (2) inform weighting controls, or (3) provide data quality checks for specific retail goods or sets of items (i.e., channels). Within specific channels or particular types of products, these data may be quite good and sufficient for weighting or assessment purposes. Retail data on item, price, and quantity could be obtained via two avenues: (1) directly from the retailers, with BLS serving as the data aggregator, or (2) via third-party companies that specialize in the aggregation of retail information.

Collection of these data by BLS directly from retailers has appeal in that the methods can be clear and well defined and the agency can exert direct control over the operation, ensuring suitable levels of quality and standardization. There are several downsides, however, to such an approach. First, it would require development of a considerable infrastructure, including retailer sampling, recruitment, regular data capture, considerable data cleaning/processing, and verification. The data requested from retailers would not be “plug-and-play” because retailers have their own (often unique) methods, units, formats, time frames, coding schemes, and standards of quality for the information they retain for business purposes. It is often the role of the data aggregator (in this case, BLS) to assume the responsibilities of translating collected data into a standardized useful form. The associated logistics and costs would likely be substantial. Undertaking such an endeavor not only would be expensive and time-consuming, but also would introduce an incremental data collection system with its own set of issues and problems, while not eliminating or even meaningfully reducing some of the current issues experienced with the CE survey approach.

Alternatively, several third-party vendors exist who specialize in the collection, cleaning, and distribution of retail information. These organizations could likely provide data on type, quantity, price, distribution channel, location, and other information at a far lower cost relative to that associated with BLS collecting the same data on its own. Unfortunately, these vendors vary considerably in the retail products and channels they

cover, as well as their geographic reach, methods for collecting information, and data completeness. Additionally, a number of potential problems can limit the utility of the data regardless of the vendor, including but not limited to:

- missing or inaccurate (often due to imputation) information;
- use of nonprobability samples, leaving in question the statistical properties of the data;
- questionable or undocumented collection and aggregation methodologies; and
- discrepancies between aggregated data and those collected by BLS, which can be attributed to a variety of sampling, collection, and aggregation differences.

An unexpected sudden change in methodology or wholesale cessation of data collection by the vendor could leave a critical gap in the CE estimation process.

Aggregate retail data are likely of best use for the CE, therefore, if information about specific products or channels is purchased from third-party vendors, not collected and aggregated directly by BLS. A thorough vetting process for such data before use is important so that the limitations of the data are well understood.

Use of alternative microdata. The previous section addressed administrative data sources from retailers. But microlevel data at the respondent level could play a role in reducing cost and/or increasing accuracy in the CES. New types of electronic data—financial records, budgeting software, store loyalty card information—may also be captured and utilized at the household or respondent level. Such data likely have the greatest utility in enhancing recall during the CE survey, but may also in some instances serve to replace survey elements, thereby reducing respondent burden, recall (measurement) error, and data collection costs.

These data may be used in one of two ways by the CE: (1) as memory cues, i.e., tools to aid respondents in their reporting of purchases that they may have otherwise forgotten and/or misreported; or (2) as data to be extracted and used *in place of* self-reported information.

As memory-jogging devices, microdata records could be printed or reviewed on a computer screen at the time the interview is conducted. Respondents would be encouraged to review their records to remind themselves of specifics of a purchase (date, place, item, and price). Utilizing records in this fashion would increase the time burden faced by respondents, but it could have a positive effect on data quality by reducing reporting error.

Extracting data directly from Internet banking, financial software, or loyalty card systems, on the other hand, generates a number of difficulties related to:

- *permission* from individuals to access their private records;
- *accessibility* if access is required through a third-party entity (such as a bank, credit card company, or loyalty card data repository);
- *differential coverage* because use of and access to such information is not universal and likely spread disproportionately within the population;
- *process complexity* by having to deal with multiple interfaces and backend data systems;
- *data incompatibility* when data elements from the source do not coincide with the categories and units required for the CE;
- *incongruent reference periods* that differ from CE requirements;
- *data discrepancies*, when internal illogical and/or missing microdata are encountered; and
- *operational challenges* with data extraction, cleaning, and verification. The extracted data would need to be processed along a separate path from the CE survey data, then integrated, leading to both time delays in reporting and additional infrastructure and resources.

Microdata from households or respondents may be of greatest utility to understanding the CE if these data are utilized as reminders and memory-jogging tools to enhance the survey process and reduce recall error. Some data elements may also be able to be gleaned from electronic sources and used to replace current survey items; however, this would introduce a significant new set of processes for the extraction, cleaning, verification, and integration of these data.

In sum, use of external data sources may be of some value in a redesigned CE process. The incremental benefits would need to be closely contrasted with the real costs of infrastructure, time, and resources required. If used in a targeted and judicious manner, both aggregate and microdata from such sources could be an effective means to improve overall data quality for the CE. With this section as context, the panel proffers the following suggestions and an overall recommendation (Recommendation 6-8):

- Identify from vendors and aggregate retail data sources appropriate for use in a five-year exploratory process that has key annual decision points based on experimentation and testing to establish the cost and measurement properties of adopting and incorporating

the capture of such data in place of currently collected microdata items.

- Identify specific microdata items that could be obtained by consent from administrative records (e.g., mortgage records, bank statements, grocery store club cards) that could be incorporated as memory cues in the collection of retail purchase data, with a similar five-year adoption window (including associated annual decision milestones).

Recommendation 6-8: BLS should pursue a long-term research agenda that integrates new technology and administrative data sources as part of continuous process improvement. The introduction of these elements should create reductions in data collection and processing costs, measurement error, and/or the statistical variance and complexity of the CPI estimate. The agenda should address the robustness of new technology and a cost/quality/risk trade-off of using external data.

Updating Capabilities

The contextual landscape for conducting national surveys is changing at an increasingly rapid pace. Because of this, it is no longer possible for a survey to be conducted in the same way for decades or even for a single decade at a time. Successful survey vendors respond to this environment by building an adaptable staff with complex methodological and statistical skill sets, and by continuously investigating new sample designs, survey methods, and estimation strategies that anticipate future changes.

Updating Internal Staff Capabilities

In light of this reality, agencies that sponsor and conduct surveys, such as BLS, need to build and maintain flexible and capable organizations and staff. BLS has a very capable group of statisticians and researchers on staff. However, a substantial focus on staff skills and organizational function is required in order to effectively respond to these changes and maintain (or improve) the quality of survey data and estimates. Of particular importance is to facilitate ongoing development of novel survey and statistical methods, to build the capacity for more complex estimation strategies required for today's best survey designs, and to build better bridges between researchers, operations staff, and experts in other organizations that face similar problems.

The panel offers the following suggestions, followed by Recommendation 6-9.

- *Develop a proactive capacity to identify changes in the CE survey context and implement research into novel survey methods that address emerging conditions and behaviors in respondent populations.*

The BLS organization and its staff must devote a greater share of resources to an active and responsive research program that focuses on emerging technologies and behavioral patterns, rather than today's dominant survey modes. It is important to develop the capacity to study the effects of alternative methods within the context of the actual production survey, not only to evaluate which method is most effective, but also to be able to quantify the impact of a methodological change on key estimates from the survey.

For any project, the focus will be to contribute to the next generation of survey methods *as they apply to BLS survey settings*, rather than to pursue incremental changes from past methodologies. It is of paramount importance to conceptualize and evaluate methodologies in the context of total survey error and to quantify the impact of multiple sources of error (e.g., coverage, nonresponse, measurement, and processing errors).

To accomplish the above, BLS will have to hire new staff or train existing staff in a variety of areas. Several staff members will need to focus on identifying and acquiring administrative or commercial data to supplement data from the CE surveys. In the process, an evaluation of the quality and accuracy of the administrative records will need to be performed. Given that the CE redesign will involve both pilot and large-scale field tests, staff trained in experimental design will be needed to conduct and evaluate the results of this research. Both quantitative and qualitative methods must be used. The required quantitative skills are relatively well known, but the qualitative methods will go beyond conducting focus groups and cognitive testing. They include the ability to understand the effects of changes in methodology on survey operations and data collection personnel, such as how implementation of new procedures might differ across regional offices.

- *Develop the ability to evaluate and implement more complex statistical methods for sampling, imputation, and model-assisted and model-based estimation to balance the demands of clients with minimizing survey burden and costs.*

Given the complexity of BLS surveys and the way they will need to be altered to accommodate the current survey landscape, standard design-based estimators will not suffice. BLS statistical staff will need to develop more current methodological expertise for a range of modern statistical methods. For example, working knowledge is needed of imputation meth-

ods that produce multiple values for planned or unplanned missing data from questionnaires, along with an understanding of how estimates and their standard errors can be generated by users. In addition, leveraging administrative and commercial data on expenditures will necessitate expertise in statistical methods for data linkage and integration. Estimation methods will require greater reliance on models and potentially the ability to create synthetic (fully imputed) datasets that can provide users with information to measure consumer behavior over time.

BLS staff must be hired or trained to carry out these activities. Knowledge of sampling techniques and weighting will not be enough. More expertise is needed in model-assisted and model-based estimation methods for sampling, imputation, estimation, data integration, and error modeling to generate data products, evaluate methodological research, and quantify error in estimates (including the impact of methodological changes).

- *Develop a more fluid bridge between operations, research, and expertise in other organizations.*

More flexibility will be gained if research and operations staffs have closer ties. Production staff can help think through the practical issues that might arise with a new method (e.g., what might work well, what problems could arise) and gain exposure to possible future changes in the survey well before they are called upon to implement them. Research staff may develop more effective experiments and gain an understanding of aspects of data collection they are not familiar with. However, it will be important for program staff and research staff to have the technical abilities necessary to communicate with one another. This means that the two staffs must have a basic understanding of what each will bring to the table for solving both the statistical and operational problems that are sure to arise in implementing any new CE survey design.

Other agencies have extensive expertise in areas that will be of interest to BLS as it redesigns the CES and other surveys. For example, the Census Bureau, which currently has responsibility for the CE data collection, has expertise in using administrative data to augment survey datasets and is devoting considerable energy to expanding its abilities in this area. Census staff have also conducted research in survey designs that administer partial questionnaires to each respondent. Joint research endeavors can be used to leverage expertise in these areas.

In addition, where institutional barriers prove detrimental to conducting responsive research, it may be wise to develop partnerships with survey vendors who are able to provide a quicker and more effective research service than is possible within BLS.

Recommendation 6-9: BLS should increase the size and capability of its research staff to be able to effectively respond to changes in the contextual landscape for conducting national surveys and maintain (or improve) the quality of survey data and estimates. Of particular importance is to facilitate ongoing development of novel survey and statistical methods, to build the capacity for newer model-assisted and model-based estimation strategies required for today's more complex survey designs and nonsampling error problems, and to build better bridges between researchers, operations staff, and experts in other organizations that face similar problems.

Obtaining Necessary Expertise Through Others

The development of tablet-based applications requires technical expertise that is likely not available at BLS or the Census Bureau at present. Rather than take the time to develop that expertise in-house, the panel urges BLS to pursue outside expertise to speed up the development and evaluation of tablet-based applications. This will require detailed knowledge of the development environment (e.g., Android, Apple iOS) and familiarity with data collection tools such as those being envisioned for the CE. Access to design and usability expertise will also be critical for the successful development of such applications, which will likely require close collaboration with BLS subject-matter staff. But relying on in-house expertise to develop the apps would likely result in development delays and possibly suboptimal designs. If BLS decides to pursue the tablet path to redesign, developing requirements for such contract work and getting outside experts engaged as soon as possible will be critical to the success of the redesign.

Recommendation 6-10: BLS should seek to engage outside experts and organizations with experience in combining the development of tablet computer applications along with appropriate survey methods in developing such applications.

Targeted Research Needed for Successful Implementation of Design Elements

The panel views the CE redesign not as one major effort, but as an ongoing process to continually address changes in the population and in survey methods. However, it is important to set priorities for aspects of the design that need the most immediate attention to achieve basic change. The CE survey methods group at BLS has undertaken many important projects to better understand survey errors and identify design features that can help reduce these errors. The panel's objective is to provide guidance on

important areas for further research and assist in their prioritization. Below, the pathways for further research are divided into those needed to inform improvements to the surveys and those needed to inform the general design that has been suggested.

Research Needed to Support the CE Redesign

The Bureau of Labor Statistics will need to conduct research to support and inform the redesign and its implementation. The panel recommends the use of tablet technology as an important new technology for collection of expenditure data, and there are numerous aspects of its implementation that require evaluation. The panel also recommends the collection of fewer, less detailed, expenditure categories in two of the prototypes, which requires evaluation of how this structure can be used to compute the CPI and how to best collect these data. The panel also recommends research on several other promising areas that may lead to further improvements of the survey to reduce burden and help obtain better quality data, presented in a separate subsection

This is by no means a comprehensive list of research areas, but an identification of several areas that need additional research, related to implementation of the proposed designs.

Use of a tablet device. All three of the proposed designs recommend the use of a tablet device. There are numerous potential benefits in using a tablet, yet they depend on how a device is selected and implemented. Even the overall advantage of a tablet over a paper instrument is an assertion that needs to be evaluated rather than taken for granted.

The highest priority research is an evaluation of the tablet technology. Criteria may include the utility, interface, robustness, data storage, transmission, and cost. Related to this is an evaluation of the optimum type of interface. The panel recommends one resembling the TurboTax model, which features separate modules that can be selected in any order and provides the respondent with an option to enter information directly into a form or to do so through a structured conversational guide. Other types of interfaces are possible, such as one that more closely aligns with a typical survey questionnaire or an interface that is more like an event history calendar. Conditional on a selected interface design, experimentation with different visual design elements that provide visually appealing and easy-to-understand features will also be beneficial. Whatever the interface design, it needs to have a self-evident navigation, as in mobile apps.

Experimentation with the structure of the instrument will also be needed. For example, the instrument can be modularized by type of expenditure (e.g., food, clothing), it can have a linear structure (e.g., reporting all

expenditures for the day), or offer both options. Each approach has unique advantages and disadvantages, yet an understanding of any measurement differences is needed first.

Experimentation will be needed in determining the best way for the technology to provide help in key entry of items. Possibilities include drop-down menus, automated “completion” of a word being entered, and other options. Experimentation is needed to develop prompts to encourage forgotten expenditures.

In addition to designing an intuitive interface, a critical aspect of the design of the application is to maintain respondent engagement and to effectively motivate respondents to report all expenditures. For example, since the respondents are provided with the tablet device, it is possible to include games and utilities that improve user engagement. The interface itself can use features of “gamification”—applying psychological and attitudinal factors underlying successful games to motivational strategies to improve user engagement, such as virtual badges, points, and status levels. Experiments will be needed in order to achieve a design that attains a high level of respondent engagement, which can in turn help to collect more accurate and complete data.

The choice of a tablet device, the design of the interface, and the addition of any motivation features are all prerequisites for the successful implementation of a tablet device to collect expenditure data. Yet there are fundamental questions about the cost feasibility and ubiquitous use of the tablet technology that need to be addressed. First among these questions is what proportion of the households (consumer units, or CUs) will have the motivation and ability to use a tablet, once an intuitive interface has been constructed. The answer to this question does not necessarily affect the use of a tablet device, but it can inform the overall design and resource allocation.

Several direct comparisons to the use of a paper-and-pencil instrument will be needed to determine the desirability of using a tablet device. Such a comparison will also allow for a better understanding of measurement differences between the two modes, as a likely final design will involve the use of both modes by different sample members. Comparisons between the two modes will be needed for both cost and quality outcomes, as a trade-off is likely. Finally, much of the objective in the CE redesign lies in reduction of respondent burden; the tablet and the paper-and-pencil administration need to be compared in terms of respondent burden.

How people keep financial records. A fruitful line of research may be to gain a better understanding of how different people and households keep financial and expenditure data. This could inform the methods used to collect the expenditure data and the design of the tablet interface, as well as

identify different subgroups for which the methods need to be different. For example, some households may have electronic records of nearly all expenditures, such as on credit card and bank account statements; others may keep paper receipts for expenditures; a third group may use a combination of methods; and yet a fourth group may not maintain sufficient records in any form. Among those who keep electronic records, some may even use specialized software that serves as a single repository of expenditure data, such as tax-related software packages.

Within a household, some may rely on a single person to be responsible for all expenditure records, while other households divide this responsibility by the type of expenditure or by the person who made the expenditure. These examples are certainly an oversimplification of what is invariably a complex and multifaceted recordkeeping phenomenon, but studies are needed to improve understanding of how households and individuals within those households keep expenditure records today.

Collecting data on a reduced set of 96 expenditure categories. Much of the burden in the CE surveys stems from the data requirements imposed on the surveys. It is imperative to conduct a study to investigate designs that minimize the number of questions and that reduce burden on respondents, in order to acquire accurate data. Both Design B and Design C require this research. The instrument can be reduced in a number of ways, but at a minimum, an evaluation is needed of the impact of collecting 96 categories of expenditures instead of the more detailed 211 expenditure categories now collected. A preliminary evaluation of the impact on the CPI, for example, can be conducted using extant data.

Use of incentives. The U.S. population has become more reluctant to participate in surveys (e.g., Groves and Couper, 1998; Stussman, Dahlhamer, and Simile, 2005), and incentives can help mitigate the effect on nonresponse. Key, however, is how incentives are incorporated into the survey design, if they are included. The panel did not venture to recommend a particular design, as this choice can only be informed through experimentation. Aspects that may warrant experimental manipulation include the structure (e.g., prepaid versus promised, household versus individual), timing (e.g., prior, during, or after completion of the supported journal), form (e.g., cash, and if cash, whether it is electronic transfer), criteria for payment (e.g., a certain level of supported journal completeness), amounts, and potential use of differential incentives (e.g., lower compliance groups, based on burden such as from the number of people in the household). More detail on this topic is covered earlier in this chapter under “Guidelines for the Use of Incentives.”

Instrument development. A substantial amount of research will be needed on the instrument development. For example, experiments are needed to (1) investigate the optimum period to ask households to keep a supported journal, (2) evaluate measurement error, nonresponse rates, and error, and (3) evaluate the stability of the estimates. Other important areas for study, especially pertinent to the proposed designs, are the optimum recall period for different types of expenditures and the optimum time between interviews.

All the proposed designs include a shift to greater reliance on self-reports, yet still involve interviewer administration to some degree. Thus, it would be beneficial to experiment with interviewer- and self-administration of different types of questions.

Privacy versus open access. Each household member can input his or her own expenses, but there is a design choice in whether to allow each respondent to see the expenditures from others in the household. The panel is recommending allowing all household members to view recorded expenditures of the household. Certainly, allowing such transparency in expenditures within the household can limit duplication of expenses, but raises a number of potential issues, ranging from potential problems related to unwanted disclosure of expenditures to other members of the household to intentional underreporting of expenditures due to the lack of privacy. Research is needed to compare providing household members with complete privacy (e.g., individual login for each member and no information sharing between accounts) versus being able to see and assess the completeness of total household expenditures.

Potential impact from reducing proxy reporting of expenditures. The use of proxy reporting invariably involves error trade-offs, which need to be evaluated. Understanding whether the additional measurement error overwhelms the reduction in nonresponse error compared to not using proxy reporting can inform the use or avoidance of proxy reporting in the redesigned survey. Schaeffer (2010) provided useful guidance to BLS for evaluating the use of proxy reporting, including separate comparisons by topic, reference period, and relationship to the sample member and the proxy respondent. As Mathiowetz (2010) pointed out, validation studies of the accuracy in reporting based on self versus proxy reports for objective measures are virtually nonexistent; further research on the CE is needed. A further complication in the direct comparison is cost; proxy reporting will likely reduce the data collection costs and will also need to be evaluated against the likely higher measurement error in the proxy reports. However, the designs proposed by the panel encourage multiple responders within the household with a minimum of additional cost.

Experiments with imputation methods and other statistical approaches. An important recommendation that is also reflected in all three alternative designs, and especially Design C, is a greater integration of statistical methods into the survey design. The designs vary along this dimension with increasing reliance on statistical methods, with two designs using subsamples with more intensive methods to calibrate the rest of the collected data to reduce measurement error and provide accurate estimates for detailed types of expenditures. At least two lines of research are needed to inform the proposed designs. All designs involve the collection of data for a small number of the weeks in a year. Weighting or imputation methods will be needed to garner annual expenditure estimates at the household level. Finally, whichever data need is addressed, different statistical methods will need to be evaluated.

Evaluation of the effectiveness of using more intensive methods. The proposed designs suggest the use of more intensive methods to improve the accuracy of the data collected on all sample members, or the use of more intensive methods on a subsample in statistical adjustments for measurement error. Whether such approaches are warranted and how they are implemented depend on the effectiveness of the more intensive methods to obtain more accurate data. The effectiveness of these methods, in turn, depends on what they entail. Therefore, experimentation is needed with different designs.

Additional Research

Although not necessitated by any of the proposed designs, several avenues for additional research can prove beneficial to the CE redesign, particularly in the long run.

Experiment with other technologies to record and extract data. Many technologies can be used to help record or extract available data on expenditures, such as scanners (including bar code scanners and receipt scanners), handheld devices and smart phones with cameras, and software that can facilitate importing of statements. These technologies are rapidly evolving and, therefore, involve the risk of being outdated in terms of hardware and software. Furthermore, the way expenditure data are stored is also changing, and a challenge for any of these technologies is to be relevant for the foreseeable future. Cautious investigation of technologies is recommended, but no single technology is likely to replace the collection of survey data.

Split questionnaire design. An area in which BLS has devoted considerable attention is the potential use of split questionnaire design—a form of matrix sampling of survey questions in which several distinct forms of the

survey are constructed (thus modules are sampled, rather than questions) and respondents randomly (although not necessarily with equal probability) assigned to one survey form.

Evaluate the utility and the ability to obtain data from additional sources. Attempts can be made to retrieve expenditure data either from other sources or directly from records that the respondents have retained. These may include credit card/bank account statements, utility statements, pay stubs, and tax records. Some guidance may be obtained from other surveys such as the Residential Energy Consumption Survey (RECS) in how to obtain permission to access these records. Most of this evaluation, however, may need to be tailored to the CE.

Augment sample with wealthy households. The wealthiest households tend to be nonrespondents at a higher rate, causing substantial problems for some uses of the CE data. A potential remedy is to augment the CE with additional samples of wealthy households, such as based on IRS records or income data linked to small geographic areas. BLS has possibilities to link the CE sample of households to existing administrative data sources, such as IRS records, that can provide some better information about nonrespondents. Design C has a base survey that would easily facilitate the selection of additional higher income households in its follow-on components.

Identify and evaluate sources of auxiliary data (e.g., retailer data). Replacement of some CE data with data from other sources, such as retailer data or data from other surveys such as the RECS, is a risky expectation, but certain survey or diary data elements may be replaced or augmented using other sources of data. More likely uses of auxiliary data, due to the different error properties and reasons for their collection, are as benchmarks that can help evaluate the CE estimates and changes in the CE estimates, and to aid in sampling, post-survey adjustments, and estimation. One example would be to leverage auxiliary data (such as income) obtainable on sampled households from the Census Bureau to estimate nonresponse bias and improve nonresponse adjustments. A broad range of auxiliary data can be considered, such as IRS data, for a multitude of uses. Permission may be needed from a household to access these data, and research could explore “opt-out” permission (rather than “opt-in”) for the CE surveys, which would allow access to a household’s data unless they receive a “no access” notification. Research by Pascale (2011) and Singer, Bates, and Hoewyk (2011) on the use of administrative records in other contexts provides useful background for conducting such research. Furthermore, these data sources change over time, and investigation of such sources needs to be ongoing rather than at one point in time.

Research Specific to a Single Prototype

These topics were listed earlier in the chapter under the descriptions of the panel's three prototypes, but are repeated here so that research needs are together in one place in this report.

Design A—Detailed Expenditures Through Self Administration:

- Develop models that would estimate quarterly and annual expenditures and income at the household level from the four weeks of reported detailed data plus the data reported on larger and routine expenditures.

Design B—A Comprehensive Picture of Expenditures and Income:

- Investigate the assumption that a “bounding” interview is unnecessary to avoid telescoping and other issues.
- Investigate the accuracy and completeness of aggregated expenditures for periods up to six months and for estimates of averages (i.e., average monthly spending for gasoline) used in this prototype to construct a full set of microdata for the entire six-month period.
- Develop appropriate models to “disaggregate” aggregated expenses using data from the one-week supported journal.
- Develop methodology for a successful component that will use an intensive interview and process based on prior collation of records and financial software to achieve a budget balance for the year at the household level as described below. Extend existing research done by Fricker, Kopp, and To (2011) to fully evaluate its potential and limitations.

Design C—Dividing Tasks Among Multiple Integrated Samples:

- Research and develop models for estimation using the base survey and two waves of data collection.
- Research and develop models for imputing at the household level “smaller expense items” collected on the *detailed expenditure component* and not on the *household profile component* into the household-level dataset to complete the overall household expense profile.

Recommendation 6-11: BLS should engage in a program of targeted research on the topics listed in this report that will inform the specific redesign of the CE.

The redesign of the CE is not a static operation, and the panel anticipates a long-term need for BLS to continue to propose, test, and evaluate new data collection methods and technologies. Thus the panel recommends that BLS maintain a methods panel to allow such testing into the future.

Recommendation 6-12: BLS should fund a “methods panel” (a sample of at least 500 households) as part of the CE base, which can be used for continued testing of methods and technologies. Thus the CE would never again be in the position of maintaining a static design with evidence of decreasing quality for 40 years.

SUMMARY

The current CE design has been in place since the late 1970s, and change is needed. The uses of the CE have grown over that time, and the current program tries to meet the needs of many users. The result is that the current surveys create an undesirable level of burden, and the data suffer from a number of quality issues. The panel believes that change should begin with BLS prioritizing the breadth and detail of data currently supporting the many uses of the CE so that a new design can most efficiently and effectively target those priorities.

The panel offers three prototype designs, each of which meets the basic requirements presented in *Consumer Expenditure Survey (CE) Data Requirements* (Henderson et al., 2011). A given prototype may be a better fit than others, depending on the revised objectives of the CE. The prototypes have considerable comparability as well. They all are designed to promote an increased use of records. They all incorporate self-administration (supported from the field representative, a tablet computer, and a centralized support facility) as a mode of data collection. They all use incentives to motivate respondents.

This report provides guidance to BLS in the next steps toward redesign. It recommends that BLS produce a roadmap for redesign within six months. The report provides guidance on how to incorporate new technology, particularly the tablet computer. The redesigned CE will still be a difficult survey for respondents, and the panel recommends developing an effective program of incentives to enhance motivation. It provides guidance in doing so.

The panel understands that a redesign of the CE will require significant targeted research to develop specific procedures that are workable and most effective. The report provides an outline of the research that is needed and the panel’s suggestions on the priority of those research endeavors. The panel recommends that BLS enhance the size and capability of its in-house research program in order to carry out the targeted research but also to

meet additional challenges in future years. Finally, the panel recommends that BLS reach out to other organizations for assistance in implementing the tablet-based data collection system and the apps that will make it work smoothly.

The panel has great confidence that BLS, with its dedicated and knowledgeable staff, will be able to move forward successfully toward a new CE. We trust that this report has helped in that process. It has been a challenging opportunity to consider these issues and make recommendations.

References and Bibliography

- Agarwal, S., D. Aaronson, and E. French. 2008. *The Spending and Debt Responses to Minimum Wage Increases*. Available: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SED2008&paper_id=379.
- Agency for Healthcare Research and Quality. 2012. *Medical Expenditure Panel Survey: Survey Background*. Available: http://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp.
- Antin, J., and E.F. Churchill. 2011. *Badges in Social Media: A Social Psychological Perspective*. Presentation at the Gamification Summit, July, Chicago, IL.
- Archibald, R., and R. Gillingham. 1980. An analysis of the short-run consumer demand for gasoline using household survey data. *The Review of Economics and Statistics* 62(4):622–628.
- Archibald, R., and R. Gillingham. 1981. The distributional impact of alternative gasoline conservation policies. *The Bell Journal of Economics* 12(2):426–444.
- Attanasio, O.P., and S.J. Davis. 1996. Relative wage movements and the distribution of consumption. *Journal of Political Economy* 104(6):1,227–1,262.
- Attanasio, O.P., and G. Weber. 1995. Is consumption growth consistent with intertemporal optimization? Evidence from the Consumer Expenditure Survey. *Journal of Political Economy* 103(6):1,121–1,157.
- Attanasio, O.P., J. Banks, C. Meghir, and G. Weber. 1999. Humps and bumps in lifetime consumption. *Journal of Business & Economic Statistics* 17(1):22–35.
- Attanasio, O.P., E. Battistin, and A. Leicester. 2006. *From Micro to Macro, from Poor to Rich: Consumption and Income in the U.K. and the U.S.* Presentation at the Conference of the National Poverty Center on Consumption, Income, and the Well-Being of Families and Children, Washington, DC, May 4–5.
- Attanasio, O.P., E. Battistin, and M. Padula. 2010. *Inequality in Living Standards since 1980: Evidence from Expenditure Data*. Washington, DC: American Enterprise Institute.
- Aune, D. 2011. *Agricultural Resource Management Survey: Integrating Varied Data Needs into a Single Data Collection*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_aune.pdf.

- Bailey, J. 2011. *Nielsen Life360 Approach*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_bailey.pdf.
- Balakrishnan, P.V., S.K. Chawla, M.F. Smith, and B.P. Micholski. 1992. Mail survey response rates using a lottery prize giveaway incentive. *Journal of Direct Marketing* 6:54–59.
- Banks, J., R. Blundell, and S. Tanner. 1998. Is there a retirement-savings puzzle? *The American Economic Review* 88(4):769–788.
- Barrett, G., P. Levell, and K. Milligan. 2012. *A Comparison of Micro and Macro Expenditure Measures Across Countries Using Differing Survey Methods*. Paper prepared for the Conference on Improving the Measurement of Consumer Expenditures, sponsored by the Conference on Research in Income and the National Bureau of Economic Research, December 2011. Available: <http://www.nber.org/chapters/c12665.pdf>.
- Battistin, E. 2003. *Errors in Survey Reports of Consumption Expenditures*. Working Paper No. 0307. London, England: Institute for Fiscal Studies.
- Beatty, P. 2010. *Considerations Regarding the Use of Global Survey Questions*. Presentation at the Consumer Expenditures Survey Methods Workshop, December 8–9, Suitland, MD. Available: http://www.bls.gov/cex/methwrkshp_pap_beatty.pdf.
- Beaule, A., and F. Stafford. 2011. *Applying Event History Methods in a National Panel: The Design and Use of Event History Calendars in the Panel Study of Income Dynamics*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_stafford.pdf.
- Bea, A., B.D. Meyer, and J.X. Sullivan. 2012. *The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?* NBER Working Paper No. 18308. Cambridge, MA: National Bureau of Economic Research. Available: http://www.nber.org/papers/w18308.pdf?new_window=.
- Bensky, N., A.T. Burks, T. Yancey, C. Shuttles, and M. Link. 2010. *Contingent Incentives Effects on Survey Response*. Presentation at the American Association for Public Opinion Research Sixty-Fifth Annual Conference, May 13–16, Chicago, IL.
- Bernheim, D., J. Skinner, and S. Weinberg. 2001. What accounts for the variation in retirement wealth among U.S. households? *American Economic Review* 91(4):832–857.
- Bhattacharya, J., T. DeLeire, S. Haider, and J. Currie. 2003. Heat or eat? Cold weather shocks and nutrition in poor American families. *American Journal of Public Health* 93(7):1,149–1,154.
- Bils, M. 2010. *Some Uses of CE*. Presentation at the 2010 CE Data Users Forum, Bureau of Labor Statistics, June 21–22, Washington, DC. Available: <http://www.bls.gov/cex/duf2010bils1.pdf>.
- Borg, P.P. 2011. *The EU Harmonisation of the Household Budget Surveys*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_borg.pdf.
- Bound, J., and A. Krueger. 1991. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics* 9(1):1–24.
- Bound, J., C. Brown, and N. Mathiowetz. 2001. Measurement error in survey data. In *Handbook of Econometrics*, S.J. Heckman and E. Learner (Eds.). Amsterdam, North Holland.
- Bowie, C. 2011. *Issues in Implementing Change in a Complex Survey*. Paper presented at the Redesign Options Workshop, October 26, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pp_bowie.pdf.

- Bradburn, N. 2010. *Recall Period in Consumer Expenditure Surveys Program*. Paper presented at the Bureau of Labor Statistics Consumer Expenditure Survey Methods Workshop, December 8–9, Alexandria, VA. Available: http://www.bls.gov/cex/methwrkshp_pap_bradburn.pdf.
- Branch, E.R. 1994. The Consumer Expenditure Survey: A comparative analysis. *Monthly Labor Review* 117(12):47–55.
- Brehm, J. 1994. Stubbing our toes for a foot in the door? Prior contact, incentives, and survey response. *International Journal of Public Opinion Research* 6:45–63.
- Brennan, M., J. Hoek, and C. Astridge. 1991. The effects of monetary incentives on the response rate and cost effectiveness of a mail survey. *Journal of the Market Research Society* 33:229–241.
- Bristol, K., N. Bensky, D. Kachhi, and M. Link. 2011. *Evaluating the Impact of “Music Downloads” as Instantly Delivered Contingent Incentives*. Paper presented at the 66th Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ.
- Browning, M., T. Crossley, and G. Weber, 2003. Asking consumption questions in general purpose surveys. *Economic Journal* 113(491):F540–F567.
- Bucks, B., and K. Pence. 2006. *Do Homeowners Know Their House Values and Mortgage Terms?* Working Paper, Federal Reserve Board of Governors. Available: <http://www.federalreserve.gov/Pubs/feds/2006/200603/200603pap.pdf>.
- Bunn, J.A., and J.E. Triplett. 1983. Reconciling the CPI-U and the PCE deflator: 3rd quarter. *Monthly Labor Review* February:37–38. Available: <http://www.bls.gov/opub/mlr/1983/02/rpt1full.pdf>.
- Bureau of Economic Analysis. 2010. *Updated Summary of NIPA Methodologies*. Available: http://www.bea.gov/scb/pdf/2010/11%20November/1101_nipa-method.pdf.
- Bureau of Economic Analysis. 2011a. *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts*. Chapter 5, Personal Consumption Accounts. Available: <http://www.bea.gov/national/pdf/ch5%20PCEforposting.pdf>.
- Bureau of Economic Analysis. 2011b. Table: *Reconciliation of Percent Change in the CPI with Percent Change in the PCE Price Index*. Available: <http://www.bea.gov/iTable/iTable.cfm?reqid=12&step=3&isuri=1&1203=62>.
- Bureau of Labor Statistics. 1978. *Consumer Expenditure Survey: Integrated Diary and Interview Survey Data, 1972–73: Total Expenditures and Income for the United States and Selected Areas*. Bulletin No. 192. Washington, DC: U.S. Department of Labor.
- Bureau of Labor Statistics. 1983. *Consumer Expenditure Survey: Diary Survey, 1980–81*. Bulletin No. 2173. Washington, DC: U.S. Department of Labor.
- Bureau of Labor Statistics. 1997. Consumer expenditures and income. In *BLS Handbook of Methods*. Washington, DC: U.S. Department of Labor.
- Bureau of Labor Statistics. 2008. Chapter 16, Consumer expenditures and income. In *BLS Handbook of Methods*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/opub/hom/homch16.htm>.
- Bureau of Labor Statistics. 2009a. *Survey Source of Data for Consumer Expenditures Survey Integrated Tables, 2009*. Available: http://www.bls.gov/cex/ce_2009source.pdf.
- Bureau of Labor Statistics. 2009b. *Consumer Expenditure Survey Compared with National Health Expenditure Accounts*. Available: <http://www.bls.gov/cex/twoyear/200607/csxnhe.pdf>.
- Bureau of Labor Statistics. 2010a. *The Consumer Expenditure Survey—30 Years as a Continuous Survey*. Available: <http://www.bls.gov/cex/ceturnsthirty.htm>.
- Bureau of Labor Statistics. 2010b. *Proxy Reporting in the Consumer Expenditure Surveys Program*. Presentation at the CE Methods Workshop, December, Washington, DC. Available: <http://www.bls.gov/cex/methwrkshpproxyrpting.pdf>.

- Bureau of Labor Statistics. 2010c. *Recall Period in the Consumer Expenditure Surveys Program*. Presentation at the CE Methods Workshop, December, Washington, DC. Available: <http://www.bls.gov/cex/methwrkshprecallperiod.pdf>.
- Bureau of Labor Statistics. 2011a. *Consumer Expenditure Homepage*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/home.htm>.
- Bureau of Labor Statistics. 2011b. *Consumer Expenditure Surveys Quarterly Interview CAPI Survey Questionnaire and Survey Instructions*. Available: <http://www.bls.gov/cex/capi/2011/cecapihome.htm>.
- Bureau of Labor Statistics. 2011c. *Consumer Expenditure Diary Survey Forms and Instructions*. Available: <http://www.bls.gov/cex/csxsurveyforms.htm#diary>.
- Bureau of Labor Statistics. 2011d. *Preliminary Results from the 2011 CE Field Representatives Survey*. Internal analysis. Washington, DC: U.S. Department of Labor.
- Bureau of Labor Statistics. 2011e. *Gemini Project Vision Document*. Available: <http://www.bls.gov/cex/ovrvwgeminivision.pdf>.
- Bureau of Labor Statistics. 2011f. *Telephone Point-of-Purchase Survey (TPOPS) Questionnaire Groups and Point-of-Purchase Survey (POPS) Expenditure Categories*. Available: <http://www.bls.gov/cpi/cpiap00a.htm>.
- Bureau of Labor Statistics. 2012. Chapter 16, Consumer expenditures and income. In *BLS Handbook of Methods*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/opub/hom/homch16.htm>.
- Cantor, D. 2010. *Discussion of Plans for Designing the Recall Period for the Consumer Expenditure Interview Survey*. Presentation at the CE Methods Workshop, December, Washington, DC. Available: http://www.bls.gov/cex/methwrkshp_pap_cantor.pdf.
- Carlson, M.D. 1974. The 1972–73 Consumer Expenditure Survey. *Monthly Labor Review* 97:16–23.
- Casey, W. 2010. *CPI Requirements of CE*. Available: <http://www.bls.gov/cex/ovrvwcpirequirement.pdf>.
- Choo, S., T.Y. Lee, and P.L. Mokhtarian. 2007. Do transportation and communications tend to be substitutes, complements, or neither? U.S. consumer expenditures perspective, 1984–2002. *Transportation Research Record* 2010:123–132.
- Christian, L.M., and D.A. Dillman. 2004. The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly* 68(1):58–81.
- Church, A.H. 1993. Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57:62–79.
- Cole, N. 2011. *Design of the National Household Food Acquisition and Purchase Survey (FoodAPS)*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_cole.pdf.
- Committee on National Statistics Panel on Redesigning the Consumer Expenditure Survey. 2011. *CE Cognitive Team Draft Report*. Internal report submitted to the Panel. National Research Council, Washington, DC.
- Cowan, C.A. 2010. *Use of CE Data in the National Health Expenditure Accounts*. Presentation at the BLS Data Users Needs Forum, June 21–22, Centers for Medicare & Medicaid Services, U.S. Department of Health and Human Services. Available: <http://www.bls.gov/cex/duf2010cowan1.pdf>.
- Creech, B.J., and B.P. Steinberg. 2011. CE source selection for publication tables. In *BLS Consumer Expenditure Survey Anthology, 2011*. Available: <http://www.bls.gov/cex/anthology11/csxanthol11.pdf>.

- De Ruijter, E., J.K. Treas, and P.N. Cohen. 2005. Outsourcing the gender factory: Living arrangements and service expenditures on female and male tasks. *Social Forces* 84(1):305–322.
- Dehejia, R., T. DeLeire, and E.F.P. Luttmer. 2007. Insuring consumption and happiness through religious organizations. *Journal of Public Economics* 91(February):259–279.
- DeLeire, T., and A. Kalil. 2005. How do cohabiting couples with children spend their money? *Journal of Marriage and Family* 67(2):286–295.
- Dillman, D.A. 2007. *Mail and Internet Surveys: The Tailored Design Method* (second edition). Hoboken, NJ: Wiley.
- Dillman, D.A., J.D. Smyth, and L.M. Christian. 2009. *Internet, Mail and Mixed-mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley.
- Dubreuil, G., J. Tremblay, J. Lynch, and M. Lemire. 2011. *Redesign of the Canadian Survey of Household Spending*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_dubreuil.pdf.
- Duetsch, M. 2008. Out-of-pocket health care spending patterns of older Americans, as measured by the Consumer Expenditure Survey. Pp. 46–51 in the *BLS Consumer Expenditure Survey Anthology, 2008*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/cex/anthology08/csxanth7.pdf>.
- Edgar, J. 2010. *Respondent Record Use in the U.S. Consumer Expenditure Interview Survey*. Office of Survey Methods Research, Bureau of Labor Statistics, U.S. Department of Labor. Presented at the American Association for Public Opinion Research Sixty-Fifth Annual Conference, May 13–16, Chicago, IL. Available: <http://www.bls.gov/cex/aaporusingrecords2010edgar1.pdf>.
- Edgar, J. 2011. *2010 CE Methods Workshop Report*. Washington, DC: Office of Survey Methods Research, Bureau of Labor Statistics, U.S. Department of Labor.
- Edgar, J., J. Davis, S. Spell, R. Verlander, and G. Wetzel. 2006. *Individual Diary Feasibility Report*. Washington, DC: U.S. Census Bureau.
- Ekici, T., and L. Dunn. 2010. Credit card debt and consumption: Evidence from household-level data. *Applied Economics* 42(4):455–462.
- Eltinge, J.L., and J.M. Gonzalez. 2007. *Properties of Alternative Sample Design and Estimation Methods for the Consumer Expenditure Surveys*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/cesrvymethsgonzale4.pdf>.
- Fan, J.X., and J.R. Burton. 2005. Vehicle acquisitions: Leasing or financing? *Journal of Consumer Affairs* 39(2):237–253.
- Fields, J. 2011a. *SIPP Core and Topical Modules: Organization and Issues*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_fields1.pdf.
- Fields, J. 2011b. *Re-engineering the SIPP: Creating the SIPP-EHC*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_fields2.pdf.
- Fisher, J.D., D.S. Johnson, J.T. Marchand, T.M. Smeeding, and B.B. Torrey. 2007. No place like home: Older adults and their housing. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 62(2):S120–S128.
- Fisher, J.D., D.S. Johnson, J.T. Marchand, T.M. Smeeding, and B.B. Torrey. 2009. Identifying the poorest older Americans. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 64(6):758–766.
- Fox, E.J., and R. Sethuraman. 2006. *Retailing in the 21st Century*. New York: Springer.

- Fricker, S., B. Kopp, and N. To. 2011. *Exploring the Feasibility of Implementing a Cash-Flow Reconciliation Approach in the Consumer Expenditure Interview Survey*. NBER Working Paper No. 12671. Cambridge, MA: National Bureau of Economic Research. Available: <http://www.nber.org/chapters/c12671>.
- García-Jiménez, C.I., and A.K. Mishra. 2011. Role of ethnicity in consumption of meat products. *Applied Economics Letters* 18(7):665–669.
- Garner, T.I., G. Janini, W. Passero, L. Paszkiewicz, and M. Vendemia. 2006. The CE and the PCE: A comparison. *Monthly Labor Review* 20–46. Available: <http://www.bls.gov/opub/mlr/2006/09/art3full.pdf>.
- Garner, T.I., R. McClelland, and W. Passero. 2009. *Strengths and Weaknesses of the Consumer Expenditure Survey from a BLS Perspective*. Presentation at the National Bureau of Economic Research, Summer Institute, Conference on Research on Income and Wealth. Bureau of Labor Statistics, U.S. Department of Labor. Washington, DC. Available: <http://www.bls.gov/cex/nber2009garner1.pdf>.
- Geisen, E., A. Richards, and C. Strohm. 2011. U.S. Consumer Expenditure Records Study. Paper prepared for the U.S. Census Bureau, RTI International, Research Triangle Park, NC. Available: <http://www.bls.gov/cex/cesrvmethsgeisen1.pdf>.
- General Accounting Office. 1996. *Alternative Poverty Measures*. GAO/GGD-96-183R. Washington, DC: Government Printing Office.
- Gentleman, J. 2011. *Collection of Hierarchical Information in the Family Section of the National Health Interview Survey*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_gentleman.pdf.
- Gieseman, R. 1987. The Consumer Expenditure Survey: Quality control by comparative analysis. *Monthly Labor Review* 110(3):8–14.
- Gillingham, R., and J.S. Greenlees. 1987. The impact of direct taxes on the cost of living. *Journal of Political Economy* 95(4):775–796.
- Gillingham, R., and J.S. Greenlees. 1990. Indexing the federal tax system: A cost-of-living approach. *Journal of Business & Economic Statistics* 8(4):465–473.
- Gonzalez, J.M., and J.L. Eltinge. 2008. *Adaptive Matrix Sampling for the Consumer Expenditure Quarterly Interview Survey*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/cesrvmethsgonzale1.pdf>.
- Gonzalez, J.M., and J.L. Eltinge. 2009. *Imputation Methods for Adaptive Matrix Sampling*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/cesrvmethsgonzale2.pdf>.
- Grainger, C.A., and C.D. Kolstad. 2010. Who pays a price on carbon? *Environmental & Resource Economics* 46(3):359–376.
- Grant, C. 2007. Estimating credit constraints among U.S. households. *Oxford Economic Papers (New Series)* 59(4):583–605.
- Greenlees, J.S., W.S. Reece, and K.D. Zieschang. 1982. Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association* 77(3):251–261.
- Grootaert, C. 1986. The use of multiple diaries in a household expenditure survey in Hong Kong. *Journal of the American Statistical Association* 81(396):938–944.
- Groves, R.M. 2006. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 70(5):646–675.
- Groves, R.M., and M.P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R.M., F.J. Fowler, Jr., M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley.

- Gruber, J. 1997. The consumption smoothing benefits of unemployment insurance. *American Economic Review* 87:192–205.
- Harris, E., and K. Perese. 2010. *Estimating the Distribution of Consumption-Based Taxes with the Consumer Expenditure Survey*. Presentation at the BLS Data Users Needs Forum, Tax Analysis Division, Congressional Budget Office, Washington, DC. Available: <http://www.bls.gov/cex/duf2010harris1.pdf>.
- Hawk, W. 2011. Household spending by single persons and married couples in their twenties: A comparison. Pp. 40–46 in the *BLS Consumer Expenditure Survey Anthology, 2011*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/cex/anthology11/csxanthol11.pdf>.
- Heathcote, J., F. Perri, and G.L. Violante. 2010. Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics* 13(1):15–51.
- Henderson, S., B. Passero, J. Rogers, J. Ryan, and A. Safir. 2011. *Consumer Expenditure Survey (CE) Data Requirements*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/cedatarequirements.pdf>.
- Hong, S.H. 2007. The recent growth of the Internet and changes in household-level demand for entertainment. *Information Economics and Policy* 19(3–4):304–318.
- Horrigan, M. 2011. *BLS Charge to the CNSTAT Panel*. Presentation at the first meeting of the Panel on Redesigning the BLS Consumer Expenditure Surveys. February, National Research Council, Washington, DC.
- Horsfield, G. 2011. *Living Costs and Food Survey*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_horsfield.pdf.
- Hough, R.S. 2011. *BRDIS Reporting Tools: Facilitating the Reporting Process*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_hough.pdf.
- Houthakker, H.S., and L.D. Taylor. 1970. *Consumer Demand in the United States: Analyses and Projections*. Cambridge, MA: Harvard University Press.
- Hsieh, C.T. 2003. Do consumers react to anticipated income shocks? Evidence from the Alaska Permanent Fund. *American Economic Review* 93(1):397–405.
- James, R.N. 2009. An econometric analysis of household political giving in the USA. *Applied Economics Letters* 16(5):539–543.
- James, R.N., and D.L. Sharpe. 2007. The nature and causes of the U-shaped charitable giving profile. *Nonprofit and Voluntary Sector Quarterly* 36(2):218–238.
- Johnson, D.S., J.A. Parker, and N.S. Souleles. 2006. Household expenditure and the income tax rebates of 2001. *American Economic Review* 96(5):1,589–1,610.
- Johnson-Herring, S., and S. Krieger. 2008. Response rates in the Consumer Expenditure Survey. In *Consumer Expenditure Survey Anthology*. Available: <http://www.bls.gov/cex/anthology08/csxanth3.pdf>.
- Kaufman, P. 2007. Strong competition in food retailing despite consolidation. *Amber Waves* 5(1):4. Available: <http://www.ers.usda.gov/AmberWaves/February07/Findings/Strong.htm>.
- King, S., B. Chopova, J. Edgar, J.M. Gonzales, D. McGrath, and L. Tan. 2009. *Assessing Nonresponse Bias in the CE Interview Survey*. Available: <http://www.bls.gov/osmt/pdf/st090220.pdf>.
- Kizakevich, P. N. 2011. *Personal Diary and Survey Methodologies for Health and Environmental Data Collection*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_kizakevich.pdf.

- Kreuter, F., S. McCulloch, S. Presser, and R. Tourangeau. 2011. The effects of asking filter questions in interleaved versus grouped format. *Sociological Methods and Research* 40(1):88–104.
- Kroshus, E. 2008. Gender, marital status, and commercially prepared food expenditure. *Journal of Nutrition Education and Behavior* 40(6):355–360.
- Krueger, D., and F. Perri. 2006. Does income inequality lead to consumption inequality? Evidence and theory. *Review of Economic Studies* 73:163–193.
- Kulka, R. 2011. *Discussion of Methodological/Cognitive Issues and the Proposed Redesigns*. Presentation at the Redesign Options Workshop, October 26, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pap_kulka.pdf.
- Lee, W.L. 2010. *BLS Consumer Expenditure Survey Data Usage: IRS Experience*. Presentation at the BLS Data Users Needs Forum. Internal Revenue Service, U.S. Department of the Treasury, Washington, DC. Available: <http://www.bls.gov/cex/duf2010lee1.pdf>.
- Lepkowski, J., and M. Couper. 2002. Nonresponse in the second wave of longitudinal household surveys. In *Survey Nonresponse*, R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.). New York: Wiley.
- Levy, H., and T. DeLeire. 2008. What do people buy when they don't buy health insurance and what does that say about why they are uninsured? *Inquiry-The Journal of Health Care Organization Provision and Financing* 45(4):365–379.
- Link, M., and A. Burks. 2012. *Address-Based Sampling: Census Block Group Data Used to Define Incentive Structure*. Presentation at the 67th Annual Meeting of the American Association for Public Opinion Research, Orlando, FL.
- Lino, M. 2010. *Expenditures on Children by Families*. Presentation at BLS Data Users Needs Forum. Center for Nutrition Policy and Promotion. U.S. Department of Agriculture, Washington, DC. Available: <http://www.bls.gov/cex/duf2010lino1.pdf>.
- Mace, B.J. 1991. Full insurance in the presence of aggregate uncertainty. *Journal of Political Economy* 99(5):928–956.
- Machlin, S. 2011. *Medical Expenditure Panel Survey: Integration of Survey Components and Linkage to Other Data Sources*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC.
- Marquis, K., and J. Moore. 1990. *Measurement Errors in SIPP Program Reports. Research Report Series*. Survey Methodology #2010-01. Washington, DC: U.S. Bureau of the Census. Available: <http://www.census.gov/srd/papers/pdf/rsm2010-01.pdf>.
- Martinez-Ebers, V. 1997. Using monetary incentives with hard-to-reach populations in panel surveys. *International Journal of Public Opinion Research* 9:77–86.
- Mathiowetz, N. 2010. *Self and Proxy Reporting in the Consumer Expenditure Survey Program*. Commissioned paper for the Council of Professional Associations on Federal Statistics, Consumer Expenditure Survey Methods Workshop, Washington, DC.
- Mathiowetz, N., K. Olson, and C. Kennedy. 2011a. *Redesign Options for the Consumer Expenditure Survey*. Presentation at the Redesign Options Workshop, October 26, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pp_abtsrbirecommend.pdf.
- Mathiowetz, N., K. Olson, and C. Kennedy. 2011b. *Redesign Options for the Consumer Expenditure Survey*. Final report prepared for the National Academy of Sciences, NAS Solicitation No. DBASSE-004950-0001-031411. Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pap_abtsrbirecommend.pdf.
- McCully, C. 2011. Trends in consumer spending and personal saving, 1959–2009. *Survey of Current Business* 91(6):14–23.

- Meyer, B.D., and J.X. Sullivan. 2008. Changes in the consumption, income, and well-being of single mother headed families. *American Economic Review* 98(5):2,221–2,241.
- Meyer, B.D., and J.X. Sullivan. 2011. Viewpoint: Further results on measuring the well-being of the poor using income and consumption. *Canadian Journal of Economics* 44(1):52–87.
- Mint.com. 2011. *How To: Track Cash Spending and Checks*. Available: <http://www.mint.com/blog/how-to/track-cash-spending-and-checks/>.
- Mockovak, W., J. Edgar, and N. To. 2010. *Results from the CEQ Field Representatives' Survey*. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.
- Moran, L.R., and C.P. McCully. 2001. Trends in consumer spending, 1959–2000. *Survey of Current Business*, March.
- National Center for Education Statistics. 2012. *National Postsecondary Student Aid Study, Survey Design and Methodology*. Available: <http://nces.ed.gov/surveys/npsas/about.asp>.
- National Research Council. 1995. *Measuring Poverty*. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods. C.F. Citro and R.T. Michael (Eds.). Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- National Research Council. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes, C.L. Schultze and C. Mackie (Eds.). Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- National Research Council. 2011a. *The Future of Federal Household Surveys: Summary of a Workshop*. K. Marton and J.C. Karberg, Rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- National Research Council. 2011b. *Project Description: Redesigning the BLS Consumer Expenditure Surveys*. DBASSE-CNSTAT-10-03. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Available: <http://www8.nationalacademies.org/cp/projectview.aspx?key=49322>.
- Nelson, J.A. 1994. On testing for full insurance using consumer expenditure survey data. *Journal of Political Economy* 102(2):384–394.
- O'Brien, E. 2011. *Energy Records in Lieu of Bills and Self Reports*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_obrien.pdf.
- Olson, K. 2011. *Responsive Survey Design for the Current Expenditure Interview Survey*. Final report prepared under contract DOLBO92J13769. Department of Sociology, University of Nebraska–Lincoln.
- Olson, K., J. Smyth, and H. Wood. 2011. *Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates?* Presentation at the Household Survey Producers Workshop, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_smyth.pdf.
- Parker, J.A. 1999. The reaction of household consumption to predictable changes in Social Security taxes. *American Economic Review* 89(4):959–973.
- Parker, J.A., N.S. Souleles, and C. Carroll. 2011. *Why Panel Data Is Indispensable for Accurate Measurement of Consumption Expenditures*. Presentation at the December 2–3 NBER-CRIW Conference on Improving the Measurement of Consumption Expenditures, Washington, DC.
- Parker, J.A., N.S. Souleles, D.S. Johnson, and R. McClelland. 2011. *Consumer Spending and the Economic Stimulus Payments of 2008*. Working Paper. Available: <http://www.kellogg.northwestern.edu/research/risk/projects/Parker%20PSJM%20Feb%2018%202010.pdf>.

- Pascale, J. 2011. *Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIPP)*. Survey Methodology #2011-03, U.S. Census Bureau. Available: <http://www.census.gov/srd/papers/pdf/ssm2011-03.pdf>.
- Passero, W. 2011. *Table 1. Summary Comparison of Aggregate Consumer Expenditures (CE) and Personal Consumption Expenditures (PCE), Based on 2002 Benchmark and Restricted to the Most Comparable Categories on the Basis of Concepts Involved and Comprehensiveness, 2003–2009*. Available: <http://www.bls.gov/cex/cecomparison.htm>.
- Paulin, G. 2008. Examining expenditure patterns of young single adults in a historical context: Two recent generations compared. Pp. 52–66 in *BLS Consumer Expenditure Survey Anthology, 2008*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/cex/anthology08/csxanth8.pdf>.
- Paulin, G. 2011. How consumers used the 2008 tax rebates: Spending, saving, or paying off debt. Pp. 29–39 in *BLS Consumer Expenditure Survey Anthology, 2011*. Washington, DC: U.S. Department of Labor. Available: <http://www.bls.gov/cex/anthology11/csxanthol11.pdf>.
- Peytchev, A. 2010. *Global Versus Specific Questions for the Consumer Expenditure Survey*. Paper prepared for the Consumer Expenditures Survey Methods Workshop, December 8–9. Available: http://www.bls.gov/cex/methwrkshp_pap_peytchev.pdf.
- Reece, W. 1979. Charitable contributions: New evidence on household behavior. *American Economic Review* 69:142–151.
- Reece, W., and K. Zieschang. 1985. Consistent estimation of the impact of tax deductibility on the level of charitable contributions. *Econometrica* 53(2):271–293.
- Ryan, J. 2010. *The CE Program: Dedicated to Improvement*. Presentation at BLS Data Users Needs Forum. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/duf2010ryan1.pdf>.
- Sabelhaus, J., D. Johnson, S. Ash, D. Swanson, T. Garner, J. Greenlees, and S. Henderson. 2012. *Is the Consumer Expenditure Survey Representative by Income*. Paper prepared for the Conference on Improving the Measurement of Consumer Expenditures, sponsored by the Conference on Research in Income and Wealth and the National Bureau of Economic Research, Washington, DC, December. Available: <http://www.federalreserve.gov/pubs/feds/2012/201236/201236pap.pdf>.
- Safir, A., and K. Goldenberg. 2008. Mode effects in a survey of consumer expenditures. Pp. 4,436–4,443 in *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
- Sakshaug, J. 2011. *To Link or Not to Link? Assessing the Quality of Administrative Data for Survey Research*. Presentation at the Washington Statistical Society and the D.C. Chapter of the American Association for Public Opinion Research, August 18, Washington, DC.
- Sampson, S.D. 2008. Category killers and big-box retailing: Their historical impact on retailing in the USA. *International Journal of Retail & Distribution Management* 36(1):17–31.
- Schaeffer, N.C. 2010. *Issues in Proxy Reporting in the Consumer Expenditure Survey*. Commissioned paper for the Council of Professional Associations on Federal Statistics, Consumer Expenditure Survey Methods Workshop, Washington, DC.
- Schenker, N., and V. Parsons. 2011. *Combining Information from Multiple Surveys: Examples with NCHS Surveys*. Presentation at the Household Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_schenker.pdf.
- Short, K.S. 2010. *CE in the SPM*. Presentation at the BLS Data Users Needs Forum. U.S. Census Bureau, Washington, DC. Available: <http://www.bls.gov/cex/duf2010short1.pdf>.
- Singer, E. 2002. The use of incentives to reduce nonresponse in household surveys. In *Survey Nonresponse*, R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.). New York: Wiley.

- Singer, E., J. Van Hoewyk, N. Gebler, T. Ragunathan, and K.A. McGonagle. 1999. The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics* 15:217–230.
- Singer, E., N. Bates, and J. Van Hoewyk. 2011. *Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census*. Presentation at the Joint Statistical Meetings, American Statistical Association. Available: <http://www.amstat.org/sections/srms/proceedings/y2011/Files/400168.pdf>.
- Slesnick, D.T. 1992. Aggregate consumption and savings in the postwar United States. *Review of Economics and Statistics* 74(4):585–597.
- Slesnick, D.T. 1993. Gaining ground: Poverty in the postwar United States. *Journal of Political Economy* 101(1):1–38.
- Souleles, N.S. 1999. The response of household consumption to income tax refunds. *American Economic Review* 89(4):947–958.
- Steinberg, B., B.J. Creech, M.L. Schmidt, and P. Falwell. 2010. *Source Selection: Selecting and Evaluating America's Expenditures*. Presentation at the Joint Statistical Meetings, Section on Government Statistics, Washington, DC. Available: <http://www.bls.gov/osmr/pdf/st100150.pdf>.
- Stephens, M., Jr. 2003. “3rd of the month”: Do Social Security recipients smooth consumption between checks? *American Economic Review* 93(1):406–422.
- Stephens, M., Jr. 2008. The consumption response to predictable changes in discretionary income: Evidence from the repayment of vehicle loans. *The Review of Economics and Statistics* 90(2):241–252.
- Stewart, H., N. Blisard, and D. Jolliffe. 2003. Do income constraints inhibit spending on fruits and vegetables among low-income households? *Journal of Agricultural and Resource Economics* 28(3):465–480.
- Stinson, L., N. To, and J. Davis. 2003. Creating a user-friendly expenditure diary. *Consumer Expenditure Survey Anthology* 3–7.
- Stussman, B., J. Dahlhamer, and C. Simile. 2005. *The Effect of Interviewer Strategies on Contact and Cooperation Rates in the National Health Interview Survey*. Presentation at the Federal Committee on Statistical Methodology Conference, Washington, DC.
- Tourangeau, R., and T. Smith. 1996. Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly* 6:275–304.
- Tourangeau, R., L.J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- Tourangeau, R., M. Couper, and F. Conrad. 2007. Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly* 71:91–112.
- Triplett, J.E., and S.M. Merchant. 1973. The CPI and the PCE deflator: an econometric analysis of two price measures. *Annals of Economic and Social Measurement* 2(3). Available: <http://www.nber.org/chapters/c9904.pdf>.
- Trussell, N., and P. Lavrakas. 2004. The influence of incremental increases in token cash incentives on mail survey response: Is there an optimal amount? *Public Opinion Quarterly* 68:349–367.
- Tucker, C. 1992. The estimation of instrument effects on data quality in the consumer expenditure diary survey. *Journal of Official Statistics* 8(1):4,161.
- U.S. Census Bureau. 2010. *Computer and Internet Use in the United States: 2010*. Available: <http://www.census.gov/hhes/computer/publications/2010.html>.
- U.S. Department of Agriculture. 2010. *Expenditures on Children by Families, 2009*. M. Lino. Center for Nutrition Policy and Promotion. Publication No. 1528-2009. Washington, DC: U.S. Department of Agriculture. Available: <http://www.cnpp.usda.gov/publications/crc/crc2009.Pdf>.

- U.S. Department of Agriculture. 2011. *Food Assistance and Nutrition Programs: The National Household Food Acquisition and Purchase Survey (FoodAPS)*. Available: http://www.ers.usda.gov/Briefing/FoodNutritionAssistance/food_aps.htm.
- U.S. Department of Energy. 2011. *How Does EIA Estimate Energy Consumption and End Uses in U.S. Homes?* Available: <http://www.eia.gov/consumption/residential/methodology/2009/brief.cfm>.
- U.S. Office of Management and Budget. 2006. *Guidelines on Agency Survey and Statistical Information Collection*. Available: http://www.whitehouse.gov/sites/default/files/omb/info/reg/pmc_survey_guidance_2006.Pdf.
- Warriner, K., J. Goyder, H. Gjertsen, P. Hohner, and K. McSpurren. 1996. Charities, no; lotteries, no; cash, yes: Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly* 60:542–562.
- Weagley, R.O., and E. Huh. 2004. The impact of retirement on household leisure expenditures. *Journal of Consumer Affairs* 38(2):262–281.
- Westat, B., and K. Olson. 2010. How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly* 5:99–114.
- Westat. 2005. *A Field Test of a Multiple Diary Procedure for the Consumer Expenditure Survey*. Report to the Bureau of Labor Statistics. Rockville, MD: Westat.
- Westat. 2011a. *Data Capture Technologies and Financial Software for Collecting Consumer Expenditure Data*. Final report to the Bureau of Labor Statistics. Available: <http://www.bls.gov/cex/ceother2011westat.pdf>.
- Westat. 2011b. *Redesign Options for the Consumer Expenditure Survey*. Presentation at the Redesign Options Workshop, October 26, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pp_westatrecommend.pdf.
- Westat. 2011c. *Redesign Options for the Consumer Expenditure Survey*. Final report prepared for the National Academy of Sciences, NAS Solicitation Number DBASSE-004950-0001-031411. Washington, DC. Available: http://www.bls.gov/cex/redwrkshp_pap_westatrecommend.pdf and http://www.bls.gov/cex/redwrkshp_app_westatrecommend.pdf.
- Wine, J., and J. Riccobono. 2011. *National Postsecondary Student Aid Study*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_wine.pdf.
- Yin, W., S.A. DeVaney, and J. Stahura. 2005. Determinants of household expenditure on computer hardware and software. *Journal of Consumer Affairs* 39(2):254–275.
- Zichermann, G., and C. Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. Sebastopol, CA: O'Reilly Media.

Appendix A

Dissent and Panel Response

This dissent is intended to clarify some of the concerns of three of the panel's economist members. Enormous effort went into this panel report and is reflected in its many good elements, which we wholeheartedly endorse. For example, the report describes several prototypical surveys. Exploring these survey types to their limit is the right approach to determine the best form for a new survey, although we are concerned about the extent to which resources will be available to BLS to pursue this course.

Despite the efforts that went into the report, some important issues did not get the attention that we think they deserve. This result is not surprising given the short time frame and a mission for the panel that included convening a data producers' workshop and commissioning and examining two prototype surveys from contractors, and a set of requirements that is challenging indeed (Casey, 2010; Henderson et al., 2011). In addition, some important information only became available to the panel late in the process. The purpose of this dissent is to highlight several issues and to stress information that the panel as a whole did not have the time to fully discuss. In particular:

- The report emphasizes a greater role for diary-style data collection, which we fear will lead to lower quality data. This emphasis is supported by a statement in the report that the quality of existing interview and diary survey data cannot be compared and other statements that are equivocal about their relative merits. This statement is inconsistent with recently estimated measures of bias and precision, two standard indicators of data quality that generally

show the existing interview data to be of higher quality than the existing diary data.

In 2010, six of the eight largest categories of expenditures available in both the Interview and Diary surveys show greater under-reporting in the Diary survey (two other large categories of expenditures are not collected in the Diary survey). The weighted average of Diary totals to national income account totals for categories that line up with the national accounts is 17 percentage points lower than the ratio for the Interview survey, though this partly reflects the different coverage of the two surveys.

For the 37 categories that can be compared in the Interview and Diary, the weighted average of the coefficient of variation is 58 percent higher in the Diary, with 33 of the 37 categories subject to greater variation even after accounting for differences in sample size (Bee, Meyer, and Sullivan, 2012). The higher variability of responses from the Diary survey mean that 2.5 diary responses are needed to equal the precision of one interview response, an issue that is not reflected in any of the cost accounting in the report.

Response rate differences between the surveys and the frequency of reports of no expenditures are also consistent with higher Interview survey data quality. The recent experience of Canada is instructive: a switch from interview to diary led to 9–14 percent more under-reporting on average (Dubreuil et al., 2011). This evidence is mentioned in the report, but is not deemed to have important implications for a redesign. Additional evidence of problems with diary surveys can be found in Crossley and Winter (2012). While these findings refer to existing surveys, they raise important concerns about the collection of expenditure data that relies heavily on diary methods.

The suggested redesigns, of course, aim to improve upon existing methods, although there is not a great deal of evidence to support that the effects will be dramatic. For example, monetary incentives for respondents are emphasized, but the evidence on such incentives from randomized trials suggests small effects that are even smaller for high-income households (e.g., Clark and Mack, 2009). The differential effects by income are particularly important since one of the most significant problems with the CE is lower response and under-reporting among high-income households (Sabelhaus et al., 2012).

- We believe that the discussion of the prototypes in the report should highlight the difficulty of using modeled data, data with incongruent time periods, and short panels. The charge from the BLS included the following requirement: “The CE considers the ability to support micro-level analyses over a 12-month period to be a requirement, with data collected at more than one point in time.” The report correctly advises the BLS that setting priorities is an important task, but we believe that the implications of the report’s recommendations for this requirement are not adequately addressed.

Modeling of microdata to construct data at the household level is problematic at best, especially since one of the purposes of the household data is to facilitate the design and estimation of behavioral models. While both statistical and economic theory provide a guide for modeling, even careful and well-intentioned methods may yield results with dramatically misleading policy implications such as the gross overstatement of the economic returns to obtaining a Graduate Equivalence Degree (GED) when using imputed observations in the Current Population Survey (CPS) (Bollinger and Hirsch, 2006). Along with different reporting periods for different goods, the lack of an annual record at the household level will prevent obtaining a complete picture of household spending.

A panel that lasts only a few weeks or months is too short for most changes in expenditures to occur. Even when expenditure changes do occur, the relatively higher dispersion due to collecting data over shorter intervals (e.g., two weeks) at each point in the panel may result in the dispersion of the change in expenditure being too unwieldy to be of use for standard panel data analysis. We feel that the discussion of prototypes should have emphasized the implications of the different designs for construction of datasets at the household level since, in our view, some of the designs will not produce usable data.

- More emphasis could have been given to the need to develop ways to monitor the quality of the data. Given that concerns about data quality are motivating the redesign, having a way to better monitor and hopefully improve data quality is vital. For instance, the ability to link to outside sources would be enhanced by a change in language in the introductory letter to respondents getting preapproval for such links such as is currently used in the American Community Survey (ACS), Survey of Income and Program Participation

(SIPP), and the CPS. These links could be used to compare survey responses to administrative or other data. A second approach—the development of an intensive “gold standard” survey that could be linked to a subset of the sample—is developed in one of the prototypes. A third approach is using internal data check, such as comparisons of spending to income that would be enhanced by having spending and income for the same time period and emphasizing budget balance. Research done with the data is also an important check on the extent to which the data have sensible patterns. Some of these issues are mentioned in the report, but we believe the goals and methods of monitoring data quality merit recommendations by the panel.

- A redesigned CE could also be more informed by how survey information affects the accuracy of the CPI. Information on what categories of expenditures and expenditures by which household members appear to be under-reported could be reflected more in the designs. Ideally, the designs would reflect which categories of expenditures for which extra detail in spending would be especially helpful in reducing bias in the CPI and which other categories have price changes sufficiently correlated across different goods to indicate that current detail is more than is needed. If the structure of the CPI as a plutocratic price index is going to continue, then under-reporting by high-income households needs to be better addressed.
- Finally, the report would benefit from further exploration of economic census data and other outside sources of data to obtain detailed expenditures. Some of these sources employ crosschecks that validate the data and make it more likely to be close to the truth. These sources would be especially useful for the calculation of CPI weights and aggregate expenditure tables.

We reiterate that these comments are not meant to criticize the work of the panel, but to emphasize that it is unfinished. The CE survey is an important statistical program, and all of the panel members have their own views on what the priorities of the survey should be and how best to address these priorities. The diversity in views on this panel has been one of its greatest strengths.

Robert Gillingham
Bruce D. Meyer
Melvin Stephens

REFERENCES

- Bee, A., B.D. Meyer, and J.X. Sullivan. 2012. *The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?* NBER Working Paper No. 18308. Cambridge, MA: National Bureau of Economic Research. Available: <http://www.nber.org/papers/w18308.pdf>.
- Bollinger, C.R., and B.T. Hirsch. 2006. Match bias from earnings imputation in the Current Population Survey: The case of imperfect matching. *Journal of Labor Economics* 24(3):483–519.
- Casey, W. 2010. *CPI Requirements of CE*. Available: <http://www.bls.gov/cex/ovrvwcpirequirement.pdf>.
- Clark, S.M., and S.P. Mack. 2009. *SIPP 2008 Incentive Analysis*. Washington, DC: U.S. Census Bureau, Demographic Statistical Methods Division.
- Crossley, T.F., and J.K. Winter. 2012. *Asking Households about Expenditures: What Have We Learned?* Paper prepared for the Conference on Improving the Measurement of Consumer Expenditures sponsored by Conference on Research in Income and Wealth and the National Bureau of Economic Research. Available: http://www.nber.org/chapters/c12666.pdf?new_window=1.
- Dubreuil, G., J. Tremblay, J. Lynch, and M. Lemire. 2011. *Redesign of the Canadian Survey of Household Spending*. Presentation at the Household Survey Producers Workshop, June 1–2, Committee on National Statistics, National Research Council, Washington, DC. Available: http://www.bls.gov/cex/hhsrvywrkshp_dubreuil.pdf.
- Henderson, S., B. Passero, J. Rogers, J. Ryan, and A. Safir. 2011. *Consumer Expenditure Survey (CE) Data Requirements*. Bureau of Labor Statistics, U.S. Department of Labor. Available: <http://www.bls.gov/cex/cedatarequirements.pdf>.
- Sabelhaus, J., D. Johnson, S. Ash, D. Swanson, T. Garner, J. Greenlees, and S. Henderson. 2012. *Is the Consumer Expenditure Survey Representative by Income*. Paper prepared for the Conference on Improving the Measurement of Consumer Expenditures, sponsored by the Conference on Research in Income and Wealth and the National Bureau of Economic Research, Washington, DC, December. Available: <http://www.federalreserve.gov/pubs/feds/2012/201236/201236pap.pdf>.

PANEL'S RESPONSE TO DISSENT

The uses of the Consumer Expenditure Surveys (CE) are complex and diverse. Our panel reflected that diversity and complexity. The members came together from different academic disciplines and collectively provided high levels of expertise in economic analysis, survey measurement, data collection, sample design, and technology. All of these skills were critical to the panel's work, and each area of expertise reflected insight, ideas, and opinions important to designing high-quality consumer expenditure surveys and to preparing a consensus report.

During panel discussions, three members continually emphasized the importance of the CE for economic analysis based on longitudinal microdata. Their ideas are important and, indeed, are part of the whole that is the panel's report. These members contributed substantially to both the substance of the report and to the balance and tone in which ideas have been laid out. The report is better because of the juxtaposition of their in-

sights and ideas with those of other panel members, and we sincerely thank them for their efforts on this report. We regret that they have now decided to emphasize their personal perspective through a dissent.

A prominent theme of the dissent is the relative quality of the current CE Interview survey in comparison with the Diary survey. The two surveys were designed differently in order for each to collect certain types of expenditures. The panel demonstrates in this report that both surveys suffer from substantial data quality issues. BLS currently uses results from both surveys, publishing estimates from one or the other of them after considering the overall quality for each individual expense item.

The majority of the panel thinks that the three dissenting members, by implying that diary-style data collection invariably produces lower-quality data than recall-style data collection, have reached a premature conclusion. The majority concluded that it is neither possible nor necessary to assess which of these two current surveys would be “better” for collecting all types of expenditures. Instead, the panel kept its focus on how to use a combination of improved modes, along with newer collection technology, external data, and respondent incentives, to create a more effective CE survey in the future. We believe the report has done that well. It provides several design options, including a supported journal using a tablet computer that is distinctly different from the current Diary survey. Each option has a different mix of interview-like and supported journal components, all of which should receive serious consideration and evaluation.

Some of the other issues in the dissent—modeling, the length of reference periods for data collection, and linking to administrative data for a richer dataset—relate directly to the dissenters’ focus on creating the most useable microdata for analysis. Each of these issues is discussed in the report, both in a broad context and in relationship to creating microdata.

The report proposes a number of ideas to improve the understanding of expenditure data and their quality, including the evaluation of intensive methods to reconcile household income and expenditures, the use of extant administrative data for quality assessment, research on how households keep financial records, and the possibility to augment the sample through more intensive interviews with a subsample of wealthier households. All are discussed in some depth in the report.

The need for an incentive program is thoroughly developed in the panel’s report. Several panel members have considerable personal experience in using and evaluating incentive programs. Knowledge of the extensive scientific literature led to our recommendation that fairly substantial incentives be used in order to affect response behavior. The report has also thoroughly developed issues related to the use of outside sources of data to obtain detailed expenditures. The panel encourages BLS to continue to

research and evaluate these options, but we concluded that none of them is currently a feasible alternative given the associated risks and costs. Far from being incomplete, we believe our report, with its broad discussions, will be an important tool for the administrators and policy makers who are responsible for determining the next steps for the CE.

Appendix B

BLS Communication of Issues

ADDITIONAL COMMUNICATION TO PANEL VIA E-MAIL FROM KATHY DOWNEY ON NOVEMBER 11, 2011

This is to followup on our conference call this afternoon. As Adam and I stated, Mike Horrigan (OPLC associate commissioner), Jay Ryan (CE division chief), and John Layng (CPI division chief) were concerned about comments overheard at the 10/27 CNSTAT meeting regarding the difficulty in designing a survey that meets all of the CPI requirements. After some discussion, the program managers decided that the CE data requirements document is sufficient.

Therefore, contrary to previous direction to the panel that both the CPI Requirements of the CE (William Casey, June 17, 2010) and the CE Data Requirements (Henderson, Passero, Rogers, Ryan, Safir, May 24, 2011) collectively form the requirements for the survey, the program managers ask that the panel members treat the CE Data Requirements as the mandatory requirements for the survey. The CPI data requirements document is still helpful in terms of providing larger context for data usage, but these are not requirements that the panel's recommendations needs to meet. We hope that this relaxation of constraints provides the panel with greater flexibility in considering their recommended design changes.

COMMUNICATION OF ISSUES PRESENTED BY
MICHAEL HORRIGAN
BLS CHARGE TO THE CNSTAT PANEL
CE PROGRAM STAFF
FEBRUARY 3, 2011

1. CE mission statement:
 - a. The mission of the Consumer Expenditure Survey program (CE) is to collect, produce, and disseminate information that presents a statistical picture of consumer spending for the Consumer Price Index, government agencies, and private data users. The mission encompasses analyzing CE data to produce socio-economic studies of consumer spending and providing CE data users with assistance, education, and tools for working with the data. CE supports the mission of the Bureau of Labor Statistics, and therefore CE data must be of consistently high statistical quality, relevant, and timely, and it must protect respondent confidentiality.
2. The technical aspects of CNSTAT's task from the Summary Statement of Work in the proposal are as follows:
 - a. The National Research Council, through its Committee on National Statistics, will convene an *Expert Panel* to contribute to the planned redesign of the Consumer Expenditure (CE) Surveys by the U.S. Bureau of Labor Statistics (BLS).
 - b. The Panel will review the output of a *data users' needs forum* and a *methods workshop*, both convened by BLS.
 - c. The Panel will conduct a *household survey data producer workshop* to ascertain the experience of leading survey organizations in dealing with the types of challenges faced by the CE surveys.
 - d. The Panel will conduct a *workshop on redesign options for the CE surveys*.
 - e. The redesign options workshop will be based on *papers on design options* the Panel commissions from one or more organizations.
 - f. Based on the workshops and its deliberations, the Panel will produce a *consensus report* at the conclusion of a 24-month study with findings and recommendations for BLS to consider in determining the characteristics of the redesigned CE surveys.
3. What CE expects from the report:
 - a. The report should synthesize information gathered through the BLS data users' needs forum, BLS methods workshop, CNSTAT household survey data producer workshop, CNSTAT CE redesign

options workshop, and independent papers into multiple comprehensive design recommendations.

- i. The design recommendations should include a menu of comprehensive design options with the highest potential, not one specific all-or-nothing design.
- ii. The design recommendations should be flexible to allow for variation in program budget, staffing resources and skills, ability of the data collection contractors to implement, legal agreements to be obtained (e.g., access to other data sources), etc.
- b. The report will include recommendations about future research that needs to be done, but that is not the focus. As much as possible, the focus should be on concrete design proposals that could be implemented.
- c. It should focus on a comprehensive design, and include an approximate timeline for development, pilot testing, and implementation. This timeline should not exceed 5 years for development and pilot testing, and a new survey in the field within 10 years.
- d. In the recommendation, the Panel should focus special attention on addressing issues with the current CE surveys:
 - i. Underreporting of expenditures
 - ii. Fundamental changes in the social environment for collection of survey data
 - iii. Fundamental changes in the retail environment (e.g., online spending, automatic payments)
 - iv. The potential availability of large amounts of expenditure data from a relatively small number of intermediaries such as credit card companies
 - v. Declining response rates at the unit, wave and item levels
- e. The Panel should develop a carefully balanced evaluation of the prospective benefits, costs, and risks of their proposed design recommendations compared to the current CE surveys. The evaluation should include a consideration of the following factors:
 - i. The evaluation be based on extensive and carefully balanced evaluation of literature and industry knowledge on methodology and practice that is currently available or likely to be available in practical form in the next five years;
 - ii. Data collection technologies currently available or likely to be available in practical form with the next five years;
 - iii. Administrative record and external data sources and technologies currently available or likely to be available in practical form with the next five years; and
 - iv. The evaluation should be reflective of the tradeoff between cost and improvement on measurement error.

4. “Yes, there are two paths you can go by, but in the long run there’s still time to change the road you’re on.” (Jimmy Page and Robert Plant, *Stairway to Heaven*)
 - a. CE is pursuing two roads to the redesign:
 - i. a redesign from scratch, and
 - ii. changes within the current design
 - b. The focus of the Panel should be on the redesign from scratch (4.a.i). In doing so, BLS would like the Panel to keep the following considerations in mind:
 - i. The Panel should be aware of the research that CE is undertaking to improve the current design.
 - ii. In considering a new design options, CE is particularly interested in approaches that focus on proactive approaches to gathering expenditure data—whether they be from records, receipts, etc. or by providing respondents the ability to easily record purchases in real time. While retrieval of data from memory in a standard reactive interview is appropriate for a number of data elements, CE views a proactive data collection methodology for expenditure data as a high priority.
 - c. As mentioned, CE is currently researching or is planning to research a number of ideas for improving the current design, including a Web diary, individual diaries, streamlining the Interview survey, reducing the length of the bounding interview, double placement of diaries, reconciliation of expenditures and income/assets, etc. (4.a.ii).
5. Constraints:
 - a. Maintain same budget.
 - b. Maintain value of the survey to taxpayers and data users.
6. What we know:
 - a. CE needs to support CPI needs.
 - b. CE needs to support other data users as much as possible as long as the design to meet those needs meets the core CE mission.
 - c. What makes CE unique is the complete picture of spending, in all categories, at the household level, with household income, assets, and demographics.
7. What we don’t know:
 - a. The final level of expenditure detail needed to support CPI’s needs after redesign.
 - i. CE has a very detailed set of current technical requirements from CPI (<http://www.bls.gov/cex/duf2010casey2.pdf>).

- ii. In cases where CE does not provide enough detail to meet CPI's needs, CPI adopts alternative approaches.
 1. For example, there are cases where the level of detail in the CE is not sufficient for CPI, such as for gasoline, food away from home, and medical care procedures.
 2. Also, there are cases where the CE sample size is not sufficient for CPI's purposes such as in calculating Entry Level Index selection probabilities at the PSU level or calculating base period weights using annual calendar periods.
- iii. CE is currently looking anew at its own data requirements and in that process will attempt to clearly state where it can and cannot meet CPI's needs in terms of CPI's current detailed technical requirements. A report will be completed by the end of April, in advance of the award of the contract for the redesign option RFP.
- iv. As the redesign process develops it is critical that ongoing dialog be maintained between CE and CPI in terms of how the redesign options would affect/change the CPI's current detailed technical requirements.
- v. In particular, CPI will need to make assessments as to the efficacy of the inputs received from CE, along with possible alternative approaches, to meet its technical requirements.
- vi. BLS views this dialog as an iterative process that must accompany the evaluation of redesign options.
- b. What importance should we place on possible future CPI information needs that could be provided by a redesigned CE?
 - i. Rob Cage's presentation, along with the document in 7.a.i above, will outline some possible future CPI information needs that could be provided possibly by a redesigned CE.
 - ii. Please note: These possible future CPI information needs are not requirements of the redesigned CE. CE views these future information needs as ones to be evaluated in terms of the following prioritized goals:
 1. Does the redesign meet the data needs of CE?
 2. Does the redesign meet the current requirements of CPI, an assessment of which includes an evaluation by CPI of the efficacy of alternative approaches in the cases where the redesigned CE does not meet its current technical detailed requirements?
 3. Within the framework of the redesigned CE, is there sufficient flexibility, especially with respect to time and cognitive burden, to collect additional data from respondents that could meet possible future information needs of CPI?

- c. Possible administrative data sources that could be used to replace some of the data CE collects, or could be used to model data.
 - d. All of the feasible technological solutions for data collection.
 - e. Data users' reaction to collecting less than the complete picture of spending and using more imputed/modeled data to create that missing data.
 - i. That is, would they find it acceptable to collect fewer data, either as part of a multiple matrix design, or because there are some expenses we won't collect, either because they are too hard to collect (like tolls on trips) or because they are such a small percentage of total spending (like reading materials)?
 - ii. Whether an approach to impute/model for a much larger amount of missing data is feasible depends on the reaction of data users and issues related to staffing and implementing a much larger statistical modeling system into production.
 - iii. Would a split sample and data collection design be feasible—one that is based on a smaller sample for which all expenditures are collected and a larger sample that takes advantage of matrix sampling and greatly reduces the burden of any given interview.
8. Consensus on the design so far:
- a. CE needs to publish a complete picture of spending, but we do not need to collect all of those data directly from respondents.
 - b. To reduce burden and improve data quality, CE is interested in moving away from a retrospective recall-based design to one that is more proactive.
 - i. The current Interview design calls for collecting almost all categories of spending from all households (the Diary is used to collect some small frequently purchased items, food, and clothing). For the most part, this collection is done through a three-month recall.
 - ii. The proposed design should not be based on a retrospective recall survey, but instead should focus on features that are proactive in collecting information from respondents or other sources. These design elements would be fundamentally different from those of the current CE surveys, and potentially include innovative features such as the use of mobile devices (e.g., smart phones, PDAs, tablets), financial software, electronic purchase records, receipt scanning, and auxiliary data.
 - iii. Retrospective recall may be incorporated into the proposed design as a method of "filling in gaps" or collecting information not otherwise provided.

- c. The constraint of maintaining the current budget needs to be considered, particularly since moving data capture from a respondent recall-based approach to one involving greater use of technology and data extraction from receipts, scanners, and administrative sources has the potential to increase collection costs.
- d. The CE program produces two main data products: published tables and microdata files. The current design is based on the idea that two surveys are needed to get a complete picture of spending—the Interview for large or regular purchases and the Diary for small or difficult-to-recall items. In the CE production process, data from the Interview and Diary are integrated at an aggregate level for publication tables; they are not integrated for the microdata files. (It may be possible to create synthetic households from the two surveys at the micro-level, but the CE program has not attempted this due to the difficulty of the project, limited resources, and the fact that historically the microdata were viewed as secondary to the publications.) This presents a problem for microdata users and means that usually they will use data from one survey or the other, but not both. With the possibility that a redesigned CE may capture data from many sources (e.g., scanners, receipts, diaries, recall interviews, administrative sources, etc.) this problem may be exacerbated. It is important that the CE program continue to make available quality microdata files to the public. These data files may include synthetic data that account for missing data in order to give a complete picture of spending at the household level. **The redesigned CE must allow for a straightforward integration of the various data sources into one complete picture of spending at the microdata level.** (Note: this is a new requirement which is not met by the current design/processing system.) As with any work that involves imputation and synthetic data, practical implementation of this approach will require a complex balance of multiple factors, including (a) implicit or explicit modeling assumptions; (b) the extent to which those assumptions are consistent with the data for specific subpopulations and expenditure types; (c) bias and variance effects arising from (a) and (b); (d) costs and complexity for the statistical agency; and (e) costs and complexity for the final data user.

Appendix C

Uses of the CE in Administering Federal Programs: Debriefing of Program Staff

The Panel on Redesigning the BLS Consumer Expenditure Surveys was particularly interested in understanding the use of the CE in administering federal programs to ensure that a redesign would continue to meet those needs. The panel spoke directly with several staff economists whose agencies use the CE data in administering their programs. The purpose of these conversations was to better understand the agency's use of the CE data, including which data were most important and how they were used. Officials were sent a letter requesting a telephone conversation.

A summary of discussions with staff at the Federal Reserve Board, Centers for Medicare & Medicaid Services of the U.S. Department of Health and Human Services, Economic Research Service of the U.S. Department of Agriculture, and Bureau of Economic Analysis appears below. These staff members spoke about their own experiences and use of the CE and were not providing an official response for their respective agencies.

Mary Kate Catlin, Economist, Centers for Medicare & Medicaid Services

Cathy A. Cowan, Economist, Office of the Actuary/National Health Statistics Group

Kenneth Hanson, Economist, Economic Research Service

Arnold Katz, Bureau of Economic Analysis

Geng Li, Economist, Federal Reserve Board

Marshall Reinsdorf, Bureau of Economic Analysis

Question 1: Do you need total expenditures for each household, as opposed to some expenditure categories collected from one set of households, and other categories collected from another set of households?

Federal Reserve: Dr. Li's use of the data would be severely impacted by this suggestion. He needs the big picture, how expenditures are related to a balance sheet. He does not know how to most efficiently use the data in this scenario. However, if the CE kept all sections for all households, but did global questions for some sections and detailed questions for others, that might work.

Centers for Medicare & Medicaid Services: Primarily use the health care section as well as income, with possible use of taxes for analysis.

Economic Research Service: Collecting some sections from some households and other sections from other households would make Dr. Hanson a little nervous about being able to do analysis of household food expenditures in the context of expenditures on other categories of goods and services (complete demand system). This survey strategy could work if enough households have at least one question per category, and then go into detail for a subset of sections for different households. He urged trying to get a large enough sample of detailed questions by category so that the analyst can distinguish different types of households, perhaps by household size and income.

Bureau of Economic Analysis: BEA worries that such a plan would eliminate the possibility of looking at substitution effects. However, the problem of underreporting of certain kinds of expenditures is so severe that a strategy of asking each respondent about only a subset of expenditures should be considered. By having sets of overlapping modules in which different respondents report on different combinations of expenditure categories, researchers who are willing to make some assumptions should still be able to study substitution effects.

Question 2: Is the panel nature of the survey important, i.e., do you need to follow households quarterly or is an annual number workable? Related, would fewer collections per household work (e.g., three times per year or twice a year)?

Federal Reserve: The strong preference is for a panel-nature survey, and even a longer panel would be extremely useful for Dr. Li's work. A modified design where the visits are every six months over two years (same number of collections) has some appeal for Dr. Li.

Centers for Medicare & Medicaid Services: The panel nature is not important, as they deal with annual numbers.

Economic Research Service: Dr. Hanson's work needs primarily an annual number. The quarterly aspect is not that important.

Bureau of Economic Analysis: Less frequent collection would end up with more boundary problems, mainly because respondents would have trouble recalling expenditures over a longer period. BEA uses quarterly data when they can get it, annual when they cannot.

Question 3: Do you need a complete financial profile for each household, e.g., income, assets, etc., in addition to expenditures?

Federal Reserve: Yes, in fact, Dr. Li and his colleagues are working with BLS to enhance the collection of assets and liabilities. This may mean a reduction in the number of questions on this, but also more accuracy and precision. Dr. Li believes that some of the questions currently being asked should be revised to match the state of household finance. He also needs income to get the complete picture.

Centers for Medicare & Medicaid Services: They do not need great financial details, but staff did express that assets are equally important to income because older people tend to have less income and more assets.

Economic Research Service: Asset data are not used, but better income data would be nice. For example, low-income households often spend more than their income, so something is missed (e.g., alimony, informal income).

Bureau of Economic Analysis: In the final survey for a household, it would be nice to have the complete picture. But the big need for BEA is for interest rates paid on big loans. Assets and liabilities would be nice.

Question 4: How detailed do you need expenses? Would global expenditures work, or if not, what level of detail do you need?

Federal Reserve: The need is more for high-level aggregate data rather than a lot of detail.

Centers for Medicare & Medicaid Services: Staff listed nine major services areas for which they need data. This is quite detailed, but not as detailed as this section gets, e.g., they don't need data on Blue Cross

Blue Shield (hospital stays, physicians, dentists, drugs, durables, health insurance, etc.).

Economic Research Service: How BLS summarizes the food data can be frustrating, and Dr. Hanson has particular concerns about the consistency of how food data are aggregated over time. He believes there is more detail in the diary than is being summarized, e.g., whole grains versus processed grains. He wants the nutritional aspects of the food to come through.

Bureau of Economic Analysis: For the most part, global expenditures would work, as long as they are accurate. For some modules, BEA would like detail in order to fill in data gaps from other sources. BEA needs to estimate a savings rate.

Question 5: Would other sources of data work for you if the CE data were not available?

Federal Reserve: The CE is unique and cannot be replaced. The PSID does not have the wealth of expenditure data and certainly not the frequency needed by this work.

Centers for Medicare & Medicaid Services: CMS uses MEPS, MCBS, and the Census Service Annual Survey.

Economic Research Service: Dr. Hanson mentioned the USDA National Household Food Acquisition and Purchase Survey (FOODAPS), conducted by Mathematica Policy Research. This survey gives good information, but it has only been funded for its current cycle, and in any best case may be a survey conducted once every five years. The annual nature of the CE is important to Dr. Hanson.

Bureau of Economic Analysis: BEA uses hundreds of sources of data. They use the CE to fill in data gaps. For example, a rental survey conducted by the Census was discontinued in 2001, and the BEA uses CE data to try to substitute for it. The need is, for example, for rent paid for single-family homes. They can get some indication of this from the CE microdata, but it's not perfect.

Dr. Katz took great pains to explain the BEA's use of many sources of data. He noted that the BEA is responsible for estimating macro-national totals such as Gross Domestic Product. To do this, they take data from where they can get it.

BEA uses CE data when the agency cannot get better data from other sources, in other words, to fill in the gaps.

BEA goes to more accurate sources of data when it can. Data on big-ticket items are more accurate from the producers, e.g., car sales best come from the auto manufacturers. But for services, these data are more likely to come from household surveys. Often, a survey such as CE is the only source of these kinds of data.

Question 6: Do you have issues with data quality of the current collection methods?

Federal Reserve: For a new user of the data, it is very hard to understand the imputations and allocations. This needs to be better explained and illustrated. There are a number of issues with internal consistency in the units from quarter to quarter. Dr. Li cited the example of a 39-year-old person who, because of an imputation in a subsequent quarter, is all of a sudden 49 years old. Dr. Li also believes that certain sections have more issues than others. For example, mortgage data have greater deficiencies. He would like internal consistency to be improved.

Centers for Medicare & Medicaid Services: Sample size is an issue since CMS produces state-level data and creates estimates for year-to-year changes. There are some level differences when compared to other data sources, but this did not seem to be a big concern.

Economic Research Service: As mentioned above, Dr. Hanson has concerns about the processing of the food data when BLS summarizes the data. He also mentioned a slide from his presentation at the 2010 Data Users Forum on the jumpiness of the year-to-year change in average food at home consumption; the jumpiness exceeds the variance.

Bureau of Economic Analysis: BEA is concerned when the data bounce around. They are concerned that this happens due to the change in panel composition every quarter, and not due to changes in consumer behavior. There are also large discrepancies between the breakdown of total expenditures derived from responses to the CE and expenditure patterns derived from sales reported by retailers and other businesses. BEA staff suspect that the main reason for these discrepancies is underreporting of expenditures by CE respondents. They believe that the sales data obtained from businesses are generally pretty accurate (but they do NOT provide as high a level of detail on the composition of

consumer expenditures as the CE). They also are concerned about the representativeness of the CE sample; in particular, higher income households account for a substantial share of total spending but getting them to answer a complete CE survey is difficult. The once-a-year release of data is a problem; BEA would like it more timely. This point, about increased need for timeliness, was made several times.

Question 7: Do you use microdata or are aggregate data acceptable?

Federal Reserve: The use of the data is primarily at the household (micro) level, but Dr. Li also uses some tabular aggregations.

Centers for Medicare & Medicaid Services: CMS uses both aggregate summaries and microdata.

Economic Research Service: Both are used.

Bureau of Economic Analysis: BEA uses microdata only when they have to in order to build up their own aggregates. They want to stay at the aggregate level. BEA does very little modeling with the CE data. They try to use the microdata as little as possible, but sometimes they don't get the cross tabulations they need, so they do some aggregations on their own from the microdata.

Question 8: Do you have any suggestions for improvement to the CE data collection?

Federal Reserve: The one strong suggestion is that when someone leaves the panel, Dr. Li would like to know why. Did the household move? Or is it a refusal?

Centers for Medicare & Medicaid Services: Timeliness is an issue. For example, they are working on 2010 estimates and the CE 2010 data won't be available until October. They would also like more information on household composition, especially older and younger than 65 years old.

Economic Research Service: Dr. Hanson believes that the diary has improved over the years, especially with recording food obtained through SNAP or other benefits. He believes that food-scanning technology is or would be a boon to the quality of the data, especially allowing nutritional content to be recorded alongside the price.

Bureau of Economic Analysis: There is great need for faster turn-around. This is true not only for collected data, but in general, in keeping up with an innovative economy. It can take years to get a new item onto a questionnaire, such as consumer electronics items, and by the time it gets on the questionnaire, it is out-of-date.

Strategies should be developed to reduce underreporting of irregular or infrequent expenditures. Devices to help people remember, such as seeing whether the reported expenditures plus saving add up to after-tax income, should be considered. Also, respondent fatigue due to the length of the survey contributes to underreporting, so breaking the survey into overlapping modules and having a respondent just report on two or three of the modules should be considered.

BEA uses the Census Bureau Economic Censuses, but these are done every five years. In between these collections, they need data to fill in the gaps. For example, the mix of sales at retail establishments can change greatly over five years. BEA might try to use CE data to estimate these mix changes indirectly. Sometimes they are dealing with seven-year-old data.

BEA identified several data gaps they feel that the CE could help with:

- Financial services; the specific services used and the fees paid.
- Data on interest rates of consumer loans, including length of loan. This is especially true of big items such as mortgages and car loans.
- Data on home maintenance and repairs. This is specifically not data on additions and alterations. BEA would like to know how much is spent just on keeping a home running.
- Online purchases, including automated bill payments.
- Keeping up with changes in the economy, including changes in consumer electronics.

Question 9: Are your needs ongoing, or more of an ad hoc use?

Federal Reserve: Dr. Li's uses of the data are ongoing. There is a long list of papers that could be written, and if the data could be improved (more of a panel nature, longer panel, less inconsistency), there would be even more to do with these data.

Centers for Medicare & Medicaid Services: Uses are ongoing, but there are some special studies.

Economic Research Service: Dr. Hanson's use is more of a one-time nature, focusing on the relationship of expenditures to nutrition, especially for low-income households.

Bureau of Economic Analysis: The needs are ongoing.

Appendix D

Statement of Work for CNSTAT Competitive Solicitation of Design Ideas

RFP #DBASSE-004950-0001-031411

INTRODUCTION

The Consumer Expenditure (CE) Quarterly Interview and Diary Surveys comprise a major program of the Bureau of Labor Statistics (BLS). The CE data are used to provide the market basket budget shares for one of the Nation's most important statistics, the Consumer Price Index. The CE surveys' unique and valuable data on the spending patterns of American consumers are used in a multitude of ways, including a new Supplemental Poverty Measure, IRS sales tax information, and economic research.

The design of the CE surveys must be updated periodically to align their methodology with changes in society, technology, consumer products, and consumer spending methods on survey estimates. Without such updates, the CE surveys will not be able to continue to fulfill their mission of producing high-quality expenditure estimates in support of their critical uses. All household surveys today face well-known challenges that include increasingly busy respondents, confidentiality and privacy concerns, many competing surveys, controlled-access residences, and non-English speaking households. In addition, the CE surveys face a unique challenge as the phenomena the surveys seek to measure have changed over the past 30 years, and continue to do so. Although the CE has made a number of improvements to the survey design, such as transitioning to computer-assisted personal interviewing (CAPI), it has not implemented changes within a systematic, comprehensive framework to address this, and other, challenges.

The CE surveys are faced with multiple issues that directly impact the quality of the data collected. These issues, presented in order of importance below, include (a) evidence of measurement error, most importantly underreporting, in the survey data, (b) changes in the survey environment due to both new technology and consumer behaviors, and (c) the need for greater flexibility in the mode of data collection. (See “Background Material” on p. 227 for links to more information.)

Measurement Error. Underreporting in the CE is evidenced by a growing deviation from other data sources and by the results of several studies. Since 1984, the ratios of aggregate expenditure estimates from the CE compared with Personal Consumption Expenditures (PCE) data from the National Accounts have shown a declining trend for many expense categories. Internal methodological studies and the 2008 CE Program Review had similar findings, providing further evidence of a growing concern about the quality of reported data. Underreporting in the CE may result from respondent burden due to survey length and complexity, panel or questionnaire conditioning, increasing telephone administration of a survey originally designed for personal visit interviews, proxy reporting by a single household member, recall effects stemming from a 3-month reference period, and/or other causes.

Changes in the Survey Environment. To remain effective, the CE surveys must adapt to changes in purchasing behaviors. For example, respondents may purchase a variety of types of items in a single large store, such as Costco or Walmart, rather than buying a single type of item from a single store. The topic-specific design of the current survey instrument may not best aid respondent recall or reporting of this type of buying. It is equally unknown whether the current survey instrument is effective for capturing on-line purchases or automatic payments. Questionnaire design and data collection methods may have to be adapted to better account for these issues. Parallel to the changes in consumer behaviors, a transformation has been occurring in the technology available for collecting data. The availability of new technology and software, such as a Web diary, a portable digital assistant (PDA) instrument, or financial software sheets, offers new opportunities to collect data. Finally, administrative sources of expenditure data (such as transaction databases built from credit/debit and from loyalty card use) now exist that provide potential alternatives to survey data.

Flexible Mode of Data Collection. One size does not fit all, and the CE needs greater flexibility in data collection modes for two main reasons: to be responsive to the needs of respondents and to allow faster imple-

mentation of changes to collection instruments. Regarding the first, a multimode design would allow data collection to be tailored to the needs and preferences of the respondent. For example, some respondents have very little time, others have difficulty keeping a diary, and still others do not want an interviewer in their home. Each respondent would have different ways of optimally reporting their data. The second motivation for greater flexibility is to allow for design changes to be made responsively as society or technology changes. Currently making changes to the CAPI instrument requires considerable lead time, and even minor changes often impact the whole instrument. A survey design utilizing multiple modes, modules, and/or technologies may allow for changes to be made to a single mode without disrupting others.

National Academy of Sciences

The Bureau of Labor Statistics has undertaken a multiyear process to develop and implement a redesign of the CE surveys to address the above issues. As part of this process BLS has contracted with the National Academy of Sciences Committee on National Statistics (CNSTAT) to conduct an independent review of the design options available, and to make specific recommendations for redesign. CNSTAT has organized a panel of experts (subsequently referred to as the panel) from across the country to carry out that task. (Panel membership is presented in Appendix A.) The panel is seeking proposals for design options and a discussion of the relative merits of the options, particularly from organizations with experience in designing complex data collection methods. It is in this context that the panel initiated this Request for Proposal through the National Academy of Sciences. A subcommittee of the panel will serve as the selection panel for this contract.

Description, Scope, and Primary Tasks

The Contractor will produce a research report laying out a design (including research studies to evaluate the proposed design) to collect information required by the primary users of the current CE. The design should be robust to recent and potential changes in the data collection and retail environment discussed above. The proposed design should focus on features that are proactive in collecting information such as the use of scanners, PDAs, or external data sources. A recall survey would be used only to supplement information not otherwise available. Thus, the proposed design would be fundamentally different from that of the current Quarterly CE survey. NAS anticipates that a team approach may be needed to meet the conditions of this contract, and the Contractor may consult with experts outside its own organization.

The report will (1) summarize the Contractor's evaluation that led them to reach the final proposal, (2) detail the proposed design, and (3) provide recommendations for evaluating the proposal in a CE-specific context. Each of these elements is specified below.

1. The Contractor will develop a carefully balanced evaluation of the prospective benefits, costs and risks of their proposed methodological and technological options compared to the current CE surveys. This evaluation should be
 - developed and presented within the context defined by the primary user needs identified in the background material.
 - based on extensive and carefully balanced review of:
 - literature and industry knowledge on household and establishment survey methodology and practice that is currently available or likely to be available in practical form in the next five years.
 - data collection technologies and external sources of expenditure data currently available or likely to be available in practical form with the next five years. Examples include, but are not limited to: bar code or receipt scanners, audio diaries, PDA instruments, and retail "loyalty card" or transaction-processor databases.
 - reflective of the tradeoff between cost and improvement on measurement error in terms of external benchmarks (e.g., CE to PCE comparisons) and other standard evaluations of data quality (e.g., subgroup comparisons).
2. Based on results from the evaluation described above, the report should detail a comprehensive proposal for a survey design, and/or other data acquisition process, which collects the required information. The proposal should describe all relevant elements for the development and implementation of the proposed design, including, but not limited to:
 - data sources;
 - sample design, including size, weighting and precision implications;
 - data collection mode(s) and/or description of alternative acquisition;
 - data collection procedures;
 - estimates of the cost requirements for development, implementation, and ongoing data collection; and
 - estimates of time requirements for development and implementation of the design, as well as time required to collect and process data to create required estimates for publication.

In the proposed design, the Contractor should address issues with the current CE surveys:

- Underreporting of expenditures.
- Fundamental changes in the social environment for collection of survey data.
- Fundamental changes in the retail environment (e.g., online spending, automatic payments).
- The potential availability of large amounts of expenditure data from a relatively small number of intermediaries such as credit card companies.
- Declining response rates at the unit, wave, and item levels.

The proposal may include a moderate number of methodological and technological options, (e.g., a mixture of household-based and establishment/intermediary-based data collection) but should go beyond simply listing possible options. Instead the report should provide, to the extent possible, concrete and specific design elements supported by in-depth explanations and evidence from the Contractor's evaluation. In other words, any proposed design should be practical and supported by evidence to demonstrate feasibility and have a rationale for expectations of improvement to the issues associated with the CE surveys.

3. In addition to proposing a new CE design, the report should include clear recommendations for evaluating the proposal in a CE-specific context, such as a pilot study. The recommendation should include
 - the design of a study or studies to evaluate the feasibility and effectiveness of the proposed design,
 - a careful description of the ways in which the study or studies would provide practical insights into the proposed new design, and
 - estimated financial and time requirements for the recommended study or studies.

Background Material

The Contractor will review background material pertinent to the potential redesign of the CE. Many of these materials can be found at the Gemini Project website (see <http://data.bls.gov/cgi-bin/print.pl/cex/geminimaterials.htm>). These materials include, but are not limited to, the following:

- Gemini Project Vision Document.
- BLS Handbook of Methods, Chapter 16, Consumer Expenditures and Income.
- BLS Statement of CE data priorities (available May 1, 2011).
- CPI Requirements of the CE.
- CE Data Quality Definition Report.
- Papers presented at the June 2010 CE Data Users' Needs Forum.
- Papers presented at the Consumer Expenditure Survey Methods Workshop, held December 8-9, 2010 in Suitland, MD.
- Papers presented at the March 2010 Data Capture Technology Forum.

Deliverables

The following deliverables are required under this contract:

- Contractor will participate in a kickoff meeting no later than fourteen (14) days following the award of this contract.
- Contractor will produce monthly progress reports on the project.
- Contractor will produce a draft report for NAS no later than ninety (90) days after the award of this contract. Draft report may be submitted electronically.
- Contractor will revise the draft report based on NAS comments and provide 8 copies of the full report developed under this proposal and based on *Description, Scope and Primary Tasks* listed above. Final report is due 120 days from the contract award.

Contractor will present the basic information in the full report to the panel at its upcoming *Workshop on Redesign Options for the Consumer Expenditure Surveys*, scheduled for October 26–27, 2011, and participate with the panel and other members of the workshop in a discussion of design options.

Appendix E

Household Survey Producers Workshop Agenda

Household Survey Producers Workshop
June 1–2, 2011
20 F Street NW Conference Center
Washington, DC 20001

AGENDA

June 1, 2011

8:30–9:15 a.m.

Welcome and Purpose of Workshop

Connie Citro, *Director, Committee on National Statistics*

Mike Horrigan, *Associate Commissioner, Office of Prices and Living Conditions, Bureau of Labor Statistics*

Don Dillman, *Panel Chair, Washington State University*

9:15–10:45

**Session 1: Alternative Ways of Measuring
Consumer Expenditures with Special Focus on
International Examples**

Chair: Mel Stephens, *Panel Member, University of Michigan*

- Canadian Survey of Household Spending, Guylaine Dubreuil, *Statistics Canada*

- UK Living Costs and Food Survey, Giles Horsfield, *Office of National Statistics*
- Household Budget Surveys in the European Union, Peter Paul Borg, *European Union*
- Discussant: Robert Gillingham, Panel Member, *Independent Consultant*

10:45–11:00

Break

11:00 a.m.–12:30 p.m. **Session 2: Designs That Add Flexibility in Data Collection Mode**

Chair: Mark Pierzchala, Panel Member, *Independent Consultant*

- Nebraska Annual Social Indicators Survey, Jolene Smyth, *University of Nebraska*
- Business R&D Innovation Survey, Richard Hough, *Census Bureau*
- National Postsecondary Student Aid Study, Jennifer Wine, *RTI*
- Discussant: Mick Couper, Panel Member, *University of Michigan*

12:30–1:45

Working Lunch to Discuss Afternoon Sessions

1:45–3:15

Session 3: Designs That Effectively Mix Data from Multiple Surveys and/or External/Administrative Data to Produce Estimates

Chair: Bruce Meyer, Panel Member, *University of Chicago*

- Medical Expenditure Panel Survey (MEPS), Steve Machlin, *AHRQ*
- Residential Energy Consumption Survey (RECS), Eileen O'Brien, *EIA*
- Behavioral Risk Factor Surveillance System and the National Health Interview Survey: Combining Data for Small Area Estimates, Van Parsons, *NCHS*
- Discussant: Clyde Tucker, Panel Member, *Independent Consultant*

3:15–3:30

Break

3:30–5:00 **Session 4: Designs That Effectively Mix Global and Detail Information to Reduce Burden and Measurement Error**

Chair: Rob Santos, Panel Member, *Urban Institute*

- Agricultural Resource Management Survey (ARMS), Dave Aune, *NASS*
- Survey of Income and Program Participation (SIPP), Jason Fields, *Census Bureau*
- National Health Interview Survey (NHIS), Jane Gentleman, *NCHS*
- Discussant: Andy Peytchev, Panel Member, *RTI*

5:00 **Working Dinner at Art & Soul to Discuss Day's Events**

7:00 **Planned Adjournment**

June 2, 2011

8:30–10:00 a.m. **Session 5: Designs That Use “Event History” Methodology to Improve Recall and Reduce Measurement Error in Recall Surveys**

Chair: David Betson, Panel Member, *University of Notre Dame*

- Panel Survey of Income Dynamics, Frank Stafford, *University of Michigan*
- Survey of Income and Program Participation, Jason Fields, *Census Bureau*
- Discussant: Michael Schober, Panel Member, *New School for Social Research*

10:00–10:30 **Break**

10:30 a.m.–12:00 p.m. **Session 6: Diary Surveys That Effectively Utilize Technology to Facilitate Recordkeeping or Recall**

Chair: Sarah Nusser, Panel Member, *Iowa State University*

- National Household Food Acquisition and Purchase Survey (FOODAPS), Nancy Cole, *Mathematica Policy Research*

- Personal Diary and Survey Methodologies for Health and Environmental Data Collection, Paul Kizakevich, *Research Triangle Institute*
- Nielsen Life360 Program, Justin Bailey, *The Nielsen Company*
- Discussant: Michael Link, Panel Member, *Nielsen*

12:00–12:30

Closing, Wrap-up

Don Dillman, Panel Chair, *Washington State University*

12:30

Planned Adjournment

Appendix F

Redesign Options Workshop Agenda

Consumer Expenditure Survey Redesign Options Workshop
October 26, 2011
20 F Street NW Conference Center
Washington, DC 20001

AGENDA

October 26, 2011

- | | |
|----------------|--|
| 8:00–8:30 a.m. | Networking and Continental Breakfast |
| 8:30–9:00 | Welcome, Introductions and Purpose of Workshop
Don Dillman, Panel Chair, <i>Washington State University</i> |
| 9:00–10:15 | Presentation of Redesign Recommendations—Westat <ul style="list-style-type: none">• David Cantor, Team Leader• Sid Schneider• Brad Edwards• Abie Reifer |
| 10:15–10:45 | Break |

- 10:45 a.m.–12:00 p.m. **Presentation of Redesign Recommendations—
University of Wisconsin–Milwaukee,
University of Nebraska, Abt-SRBI**
- Nancy Mathiowetz, Team Leader, *University of Wisconsin–Milwaukee*
 - Kristen Olson, *University of Nebraska*
 - Courtney Kennedy, *Abt-SRBI*
- 12:00–1:00 **Working Lunch for Informal Discussion of
Redesign Ideas**
- 1:00–1:30 **Discussant: Methodological/Cognitive Issues
and the Proposed Redesigns**
- Richard Kulka, *Consultant*
- 1:30–2:00 **Discussant: Implementing Change in a Large
Ongoing Survey and the Proposed Redesigns**
- Chet Bowie, *NORC (retired from U.S. Census Bureau)*
- 2:00–2:40 **Discussant Panel: Data Users of the Consumer
Expenditure Survey**
- Clinton McCully, *Bureau of Economic Analysis*. Macrodata use by a Federal program
 - Mark Lino, *Center for Nutrition Policy and Promotion, USDA*. Microdata use by a Federal program
 - Melvin Stephens, *University of Michigan*. Microdata use by a university researcher
- 2:40–3:00 **Break**
- 3:00–4:45 **Open Discussion with Panel and Workshop
Participants**
- Michael Horrigan, *Associate Commissioner, BLS*
Discussion Leader
- 4:45–5:00 **Ending Remarks and Wrap-up**
- Don Dillman, Panel Chair, *Washington State University*
- 5:00 **Planned Adjournment**

Appendix G

Biographical Sketches of Panel Members and Staff

DON A. DILLMAN (*Chair*) is Regents professor in the Department of Sociology at Washington State University. He also serves as the deputy director for research and development in the Social and Economic Sciences Research Center at Washington State University. From 1991 to 1995, he served as the senior survey methodologist in the Office of the Director at the U.S. Census Bureau. His work at the Census Bureau resulted in his receiving the Roger Herriot Award for Innovation in Federal Statistics in 2000. He is recognized internationally as a major contributor to the development of modern mail, telephone, and Internet survey methods. Throughout his time at Washington State University, he has maintained an active research program on the improvement of survey methods and how information technologies influence rural development. He has served as investigator on more than 80 grants and contracts worth approximately \$12.5 million, and written 13 books and more than 235 other publications. He holds numerous memberships in professional organizations, including the American Association for the Advancement of Science, American Sociological Association, and American Statistical Association. He served as past president of the American Association for Public Opinion Research and the Rural Sociological Society. He has a B.A. degree in agronomy, an M.S. degree in rural sociology, and a Ph.D. degree in sociology, all from Iowa State University.

DAVID M. BETSON is an associate professor of public policy and economics in the College of Arts and Letters at the University of Notre Dame. His research has focused on the impact of tax and transfer programs on the economy and the distribution of income. Of particular interest is child sup-

port policy. He has written academic papers and consulted with numerous state governments on the development of their child support guidelines. In 2007, he was appointed to the Washington State Commission on the Review of Child Support Guidelines. He holds memberships in the Association for Public Policy Analysis and Management and the American Economic Association. He has a B.A. degree in economics and physics from Kalamazoo College, and an M.A. degree in economics and a Ph.D. degree in economics, both from the University of Wisconsin–Madison.

MICK P. COUPER is a research professor in the Survey Research Center, Institute for Social Research, at the University of Michigan, and in the Joint Program in Survey Methodology at the University of Maryland and University of Michigan and Westat. He is also a faculty associate in the Population Studies Center and the Comprehensive Cancer Center, both at the University of Michigan. His current research includes survey nonresponse, design and implementation of survey data collection, effects of technology on the survey process, and computer-assisted interviewing, including both interviewer-administered (CATI and CAPI) and self-administered (Web, audio-CASI, IVR) surveys. Many of his current projects focus on the design of Web surveys. He is the recipient of several awards, including the American Association for Public Opinion Research (AAPOR) book award (with Robert M. Groves) and the AAPOR award for innovation. He is an elected fellow of the American Statistical Association and a member of several other professional organizations including AAPOR and the European Survey Research Association (ESRA). He has an M.Soc.Sc. degree in sociology from the University of Cape Town, an M.A. degree in applied social research from the University of Michigan, and a Ph.D. degree in sociology from Rhodes University in South Africa.

ROBERT F. GILLINGHAM is an independent consultant from Potomac Falls, Virginia. Current and recent clients include the International Monetary Fund, the World Bank, and the OECD. Prior to becoming a consultant, he held several positions at the International Monetary Fund, including chief of the Expenditure Policy Division and chief of the Poverty and Social Impact Analysis Group. He also worked at the U.S. Department of the Treasury—the last 10 years as deputy assistant secretary for economic policy. He started his career at the U.S. Bureau of Labor Statistics, rising to the position of deputy associate commissioner for prices and living conditions. He is currently a member of the National Academy of Social Insurance and has previously served on the board of directors of the Western Economic Association International and as an associate editor of *Journal of Business and Economic Statistics*. He is widely published, and his articles have appeared in numerous publications, including the *Review of Economics and*

Statistics, Energy Journal, and Journal of Economic Business and Statistics. He has a B.A. degree in economics from Haverford College and a Ph.D. degree in economics from the University of Pennsylvania.

CAROL C. HOUSE (*Study Director*) is a senior program officer for the Committee on National Statistics. She retired from the National Agricultural Statistics Service (NASS), USDA, in 2010, where she was deputy administrator for programs and products and chair of the Agricultural Statistics Board. Her previous positions at NASS included associate administrator, director of research and development, and director of survey management. She has provided statistical consulting on sample surveys in China, Colombia, the Dominican Republic, and Poland. She is a fellow of the American Statistical Association and an elected member of the International Statistical Institute. Her graduate training was in mathematics at the University of Maryland.

MICHAEL W. LINK is vice president for Research Methods Center of Excellence and chief methodologist at the Nielsen Company in Atlanta, Georgia. Prior to this he was a senior survey methodologist and acting branch chief at the National Center for Chronic Disease Prevention and Health Promotion at the Centers for Disease Control and Prevention. At Nielsen, he is responsible for identifying and monitoring the implementation of the most cost-effective methodological/operational solutions in order to constantly improve the quality of research methodologies. His work at the Center for Excellence focuses on identifying problems related to, and solutions for, improving the quality of research methodologies, including analysis of existing data sources, and implementation of best practices for these research methods. He has published more than 50 peer-reviewed articles and book chapters in journals such as *Public Opinion Quarterly*, *International Journal of Public Opinion Research*, *Journal of Official Statistics*, *Survey Research Methods*, and *Field Methods*, and presented research findings at more than 150 national and international scientific conferences. He has a B.S. degree in biology from Georgia State University, and M.A. and Ph.D. degrees in political science from the University of South Carolina.

BRUCE D. MEYER is the McCormick Foundation professor of public policy in the Harris School of Public Policy Studies at the University of Chicago. Prior to this appointment, he was a professor in the Economics Department at Northwestern University, where he taught for 17 years. His current research includes studies of poverty and inequality, tax policy, welfare policy, unemployment insurance, workers' compensation, disability, the health care safety net, labor supply, and the accuracy of household surveys. Previously, he was a visiting faculty member at Harvard University,

University College London, and Princeton University. He is a research associate of the National Bureau of Economic Research and a member of the National Academy of Social Insurance and the Conference on Research on Income and Wealth. He has been associated with the Institute for Research on Poverty and is a faculty fellow at the Institute for Policy Research. He has served as an advisor to the U.S. Department of Labor, U.S. Bureau of Labor Statistics, New York State Office of Temporary and Disability Assistance, Human Resources Development Canada, Manpower Demonstration Research Corporation, and Mathematica Policy Research. He has B.A. and M.A. degrees in economics from Northwestern University and a Ph.D. degree in economics from the Massachusetts Institute of Technology.

SARAH M. NUSSER is a professor in the Department of Statistics and a faculty member in the Human Computer Interaction Graduate Program at Iowa State University. She is also a faculty member in the Ecology and Evolutionary Biology Graduate Program at Iowa State University. Prior to this she was a senior research fellow at the Bureau of Labor Statistics and a statistician at the Proctor and Gamble Company. Her research interests include using geospatial data in survey data collection and estimation, estimation methods for land cover map accuracy assessment, and sample design and measurement in surveys. Her experience includes service on the Census Advisory Committee of Professional Associations and with administrative records databases through research involving welfare program evaluation and numerous operational survey projects. She received the Board of Regents Award for Faculty Excellence at Iowa State University in 2010 and the Distinguished Achievement Award from the American Statistical Association's Section on Statistics and the Environment in 2007. She has a B.S. degree in botany from the University of Wisconsin–Madison, M.S. degrees in botany and statistics, and a Ph.D. degree in statistics from Iowa State University.

ANDY PEYTCHEV is a senior survey methodologist in the Program for Research in Survey Methodology at RTI International and instructor in the Odum Institute at the University of North Carolina, Chapel Hill. He has collaborated on various experiments on optimization of data collection designs, aimed at reduction of cost, bias, and variance of survey estimates. He has led research on the design and redesign of major government surveys and is principal investigator on grants to study methods for estimation, reduction, and correction for survey nonresponse. His work has been published in peer-reviewed journals such as *Public Opinion Quarterly*, *Journal of Official Statistics*, and *Sociological Methods & Research*. He has received recognition for his service as a founding associate editor of *Survey Practice* and is currently an associate editor of *Public Opinion*

Quarterly. He is a member of the American Statistical Association and the American Association for Public Opinion Research, for which he has served on various committees. He has a B.A. degree in marketing from Concord University, an M.S. degree in survey research and methodology from the University of Nebraska–Lincoln, and a Ph.D. degree in survey methodology from the University of Michigan.

MARK M. PIERZCHALA is a private consultant providing systems, methods, and operations consulting services on survey-taking to private and government clients. Previously, he was a senior research statistician at the National Agricultural Statistics Service, where he became well known for his knowledge of data editing systems and the data cleaning methods they represented. He has a long history in data collection and data editing systems for complex U.S. and Canadian government surveys. He is an expert in the use of the Blaise survey processing system for data collection and postcollection editing. Currently, he is assisting with the testing and documentation of the emerging Blaise Next Generation system. At Westat, he was a senior systems analyst and was head of the Blaise Services Group. He helped the Bureau of Labor Statistics and the Census Bureau implement CAPI for the Consumer Expenditure Survey. Later at Mathematica, he had major roles in establishing the Blaise system for all modes of data collection, made methodological and systems advances for multimode surveys, and was the systems lead on complex multi-instrument studies. He has a B.S. degree in mathematics from Central Michigan University and an M.S. degree in mathematical statistics from Michigan State University.

ROBERT SANTOS is the senior institute methodologist in the Statistical Methods Group at the Urban Institute in Washington, DC. Prior to this, he was executive vice president and partner of NuStats Partners, LP, a social science research firm based in Austin, Texas. He also worked at NORC at the University of Chicago and the Survey Research Center at the University of Michigan. His expertise is in sampling, survey design, and survey methods and operations. His professional credits include numerous reports and papers and leadership roles in survey research associations. He has served as a member of the Census Advisory Committee of Professional Associations and on the editorial board of *Public Opinion Quarterly*. He has held numerous elected and appointed leadership positions in both the American Statistical Association and the American Association for Public Opinion Research. He is a fellow of the American Statistical Association and a recipient of its 2006 Founder's Award for excellence in survey statistics and contributions to the statistical community. He has a B.A. degree in mathematics from Trinity University and an M.A. degree in statistics from the University of Michigan.

MICHAEL F. SCHOBER is dean and professor of psychology at the New School for Social Research in New York City. He has taught numerous graduate and undergraduate courses on subjects including human computer interaction, research methods, and psycholinguistics. His research focuses on interaction in survey interviews and the effects of new technologies on communication. He is the editor of the journal *Discourse Processes*, a position he has held for the past five years. He is widely published, and his articles have appeared in numerous publications including *Cognitive Psychology*, *Journal of Official Statistics*, and *Public Opinion Quarterly*. He holds many memberships in professional organizations, including the American Psychological Association, the American Psychological Society, and the American Association for Public Opinion Research. He has a Sc.B. degree in cognitive science from Brown University and a Ph.D. degree in psychology from Stanford University.

MELVIN STEPHENS, JR., is an associate professor of economics and associate professor of public policy at the University of Michigan and a research affiliate and faculty associate in the Institute for Social Research at the University of Michigan. He is also a research associate in labor studies at the National Bureau of Economic Research. Prior to this, he was the Raymond John Wean Foundation Career Development associate professor of economics in the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University. His research interests cover such areas as displaced workers, household consumption decisions, and aging and retirement. Currently, he is on the editorial board of *The B.E. Journal of Economic Analysis and Policy* and is a member of numerous organizations, including the American Economic Association and Society of Labor Economists. He previously served as a referee for several publications, including the *American Economic Review*, *Journal of Political Economy*, and the *Quarterly Journal of Economics* and has been a reviewer for the National Institute on Aging and the National Science Foundation, among others. His work is widely published, having appeared in such publications as the *American Economic Review*, *American Economic Journal: Applied Economics*, the *Economic Journal*, and the *Review of Economics and Statistics*. He has a B.A. degree in economics and mathematics from the University of Maryland and a Ph.D. degree in economics from the University of Michigan.

CLYDE TUCKER is a former senior survey methodologist at the U.S. Bureau of Labor Statistics (BLS) where he served for more than 25 years. While at BLS, he cochaired the Interagency Research Group, which was responsible for revising the methodology for collecting information on race and ethnicity in federal surveys. He also served as a statistical consultant

to the bipartisan Congressional Commission that assessed the impact of the Family and Medical Leave Act, and was a member of the committee overseeing the methodology of the Current Population Survey. His research interests include telephone survey design, survey nonresponse, and measurement error. He is an elected fellow of the American Statistical Association (ASA), is a past chair of the Government Statistics Section of ASA, and has held numerous positions on the AAPOR National Council. He is also a past president of the Washington Statistical Society. He is a past winner of both the Herriot Award for Innovation in Federal Statistics from the American Statistical Association and the Innovator Award from the American Association for Public Opinion Research. In election years 2004, 2006, 2008, 2010, and 2012, he headed the decision desk for CNN. He has an M.S. degree in statistics and a Ph.D. degree in political science, both from the University of Georgia.

COMMITTEE ON NATIONAL STATISTICS

The Committee on National Statistics (CNSTAT) was established in 1972 at the National Academies to improve the statistical methods and information on which public policy decisions are based. The committee carries out studies, workshops, and other activities to foster better measures and fuller understanding of the economy, the environment, public health, crime, education, immigration, poverty, welfare, and other public policy issues. It also evaluates ongoing statistical programs and tracks the statistical policy and coordinating activities of the federal government, serving a unique role at the intersection of statistics and public policy. The committee's work is supported by a consortium of federal agencies through a National Science Foundation grant.

