

## Supporting Statement

### B. Collection of Information Employing Statistical Methods

#### 1. Sampling Method

##### Consumer Units

There are approximately 122 million consumer units (CUs) in the Consumer Expenditure (CE) Survey's universe.<sup>1</sup> A CU is the unit from which we seek expenditure reports. It consists of all household members in a particular housing unit or other type of living quarters who are related by blood, marriage, adoption, or some other legal arrangement. The CU determination for unrelated persons is based on financial dependence in three expenditure categories: shelter, food, and all other expenses. Unrelated persons are considered to be separate CUs if they are responsible for paying their own expenses for at least two of these categories, and they are considered to be part of the same CU if they share expenses for at least two of these categories. Approximately 97 percent of all occupied living quarters are occupied by a single CU.

For an overview of the CE sample design and the CU selection process, please refer to the 2008 CE Anthology article, "Selecting a Sample of Households for the Consumer Expenditure Survey" by Susan King and Sylvia Johnson-Herring (Attachment O).

The following table shows estimated numbers of CUs in all 91 strata from which PSUs were selected.<sup>2</sup> (See the section below entitled "PSUs" for more information.)

#### Estimated Number of CUs in CE's 91 Strata

Stratum Code	Estimated Number of CUs in Stratum
A102	2,682,337
A103	3,076,725
A104	1,109,160
A109	3,471,120
A110	3,345,637
A111	2,887,479
A207	3,975,571
A208	2,365,042

A209	1,161,194
A210	1,276,845
A211	1,359,359
A312	2,179,256
A313	1,106,574
A316	2,286,803
A318	2,043,853
A319	1,820,983
A320	1,680,184
A321	1,017,906
A419	5,359,776
A420	1,737,241

A422	3,051,152
A423	1,540,781
A424	1,219,632
A425	1,098,491
A426	423,037
A427	154,316
A429	1,409,498
A433	1,269,841
X102	1,984,754
X104	1,153,681
X108	1,527,077
X210	1,541,319

<sup>11</sup> The number of CUs comes from dividing the Census Bureau's 2011 estimate of the number of people in the civilian non-institutional population (305 million) by the average number of people per CU (2.5).

<sup>2</sup> The number of CUs per stratum comes from allocating the nationwide total of 122 million CUs by each stratum's proportion of the nationwide population in the 2000 Census.

X212	1,031,743
X218	1,176,233
X220	1,190,474
X222	1,306,535
X224	1,867,021
X226	1,054,968
X228	818,698
X232	757,580
X336	1,717,072
X338	1,193,581
X340	1,229,480
X342	1,598,293
X344	1,251,499
X346	1,133,546
X350	961,323
X352	1,227,622
X354	1,767,779
X356	1,247,878
X358	1,750,423
X362	1,598,245
X364	1,783,151
X366	904,262
X368	1,252,451
X472	1,780,653
X474	2,009,485
X482	1,504,454
X484	1,315,058
Y102	589,789
Y104	685,725
Y206	864,751
Y208	1,028,433
Y210	857,333
Y212	1,028,146
Y314	851,877
Y316	984,162
Y318	816,244
Y320	859,416
Y322	1,016,067
Y324	813,425
Y426	553,445
Y428	505,920
Y430	623,627
Y432	603,219
Z102	387,673
Z104	603,031

Z206	890,485
Z208	654,787
Z210	805,152
Z212	956,245
Z314	810,396
Z316	769,006
Z318	814,876
Z320	1,069,877
Z322	989,211
Z324	588,576
Z426	284,458
Z428	247,159
Z430	341,514
Z432	359,851
<b>Total</b>	<b>122,000,000</b>

## Response Rates

The following table shows expected 2013 sample sizes for the Quarterly Interview Survey (CEQ) and the Diary Survey (CED) based on the target number of completed interviews and 2011 response rates.

Category	Quarterly (quarter) (1)	Diary (annual)
Total Sample Size	15,320	11,750
Total Type B and C Noninterviews (vacant, demolished, etc.):		
Number	3,064	2,350
Percent of Total Sample	20.0	20.0
Total Eligible Units	12,256	9,400
Total Type A Noninterviews		
Number	3,432	2,350
Percent of Total Eligible	28.0	25.0
Total Completed Interviews		
Number	8,825	7,050
Percent of Total Eligible	72.0	75.0

(1) This is the expected quarterly sample size for all five waves of the survey.

For more information on the calculation of response rates, please see the 2008 CE Anthology article “Response Rates in the Consumer Expenditure Survey,” by Sylvia Johnson-Herring and Sharon Krieger (Attachment P).

In 2008 CE staff conducted a nonresponse bias study to determine whether the missing data from nonrespondents generated any bias in the Interview survey’s published estimates. Their study was undertaken in response to OMB’s recent directive on the matter. They synthesized the results from four individual studies, and concluded that CE’s data are not “missing completely at random,” but that no bias is introduced into the survey’s published estimates in spite of that fact. As they said, “the results from these four studies provide a counterexample to the commonly held belief that if a survey’s data are not missing completely at random (MCAR) then its estimates are subject to nonresponse bias.” For more information, please see “Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey” (Attachment T).

## 2. Collection Methods

### Primary Sampling Units (PSUs)

The primary sampling units (PSUs) used in the CEQ and CED are small clusters of counties. The average number of counties in a PSU is approximately five. The set of sample PSUs used in the two CE surveys consist of 91 PSUs, 75 of which are also used in the Consumer Price Index (CPI). The 91 PSUs fall into four categories:

PSU	Number	Description
-----	--------	-------------

<b>“size class”</b>	<b>of PSUs</b>	
A	28	Large Metropolitan CBSAs (self-representing PSUs)
X	31	Small Metropolitan CBSAs (non-self-representing PSUs)
Y	16	Micropolitan CBSAs (non-self-representing PSUs)
Z	16	Non-CBSA Areas (non-self-representing PSUs)

The BLS selected these PSUs from a stratified sample design in which one PSU was selected from each stratum. Stratification of the non-self-representing PSUs (the X, Y, and Z PSUs) used a 5-variable model whose independent variables were latitude, longitude, latitude squared, longitude squared, and the percent of consumer units living in an urban area.

### Sampling Within PSUs

Four non-overlapping sampling frames are used to select CUs for the CE surveys. The four frames are: Unit, Area, Group Quarters (GQ), and Permit. The first three frames, Unit, Area, and GQ, are called “old construction”; and the Permit frame is called “new construction.” The Permit frame consists of housing units built after April 1, 2000.

The Unit frame covers 80% of the sample, and consists of addresses of housing units located in census blocks in areas that issue building permits, and in which a high percentage of the addresses are “complete” (they contain a street name and house number). The Area frame covers 10% of the sample, and consists of addresses located in census blocks that are not in permit-issuing areas, or where more than 4 percent of the addresses in the blocks are incomplete or missing. These addresses are mostly in rural areas. The GQ frame covers 1% of the sample, and consists of boarding houses, hotel rooms, and institutions that are found in the decennial census but are not counted as housing units. These addresses are located in census blocks that contain complete addresses and where building permits are issued. Since the census does not cover housing units built after April 1 of the census year, the list of addresses in these three frames is supplemented by a sample of building permits from the Building Permit Survey, which is conducted by the Census Bureau. The Permit frame is updated by this survey monthly and covers about 9% of the sample.

Within each PSU, a “systematic sample” of households is selected from each of the four frames. The first step in the selection process is sorting the households by variables that are correlated with their expenditures. The purpose of this is to ensure that households of every wealth level are well-represented in the sample. The first household in the systematic sample is selected from the sorted list using a random number generator. After the initial household is selected, another household is selected every “k” households down the list, where “k” is the inverse of the PSU’s probability of selection times the PSU’s sampling interval.

Each frame has different sort variables. For the Unit frame, each address is assigned to a category based on whether the address is rental or owned property. Both the rental and owned categories are subdivided into quartiles of rental and property values, which are defined uniquely for each county. These eight categories are further subdivided by whether the housing unit is vacant or occupied by 1, 2, 3, or more than 4 people, and each cell is assigned a stratification code value (see table 1).

Table 1. CE Unit Frame Stratification Code Values

Renter/Owner Quartile	Number of Occupants				
	Vacant	1 person	2 persons	3 persons	4+ persons
Renters 1 <sup>st</sup> Quartile	10	11	12	13	14
Renters 2 <sup>nd</sup> Quartile	25	24	23	22	21
Owners 1 <sup>st</sup> Quartile	30	31	32	33	34
Owners 2 <sup>nd</sup> Quartile	45	44	43	42	41
Renters 3 <sup>rd</sup> Quartile	50	51	52	53	54
Renters 4 <sup>th</sup> Quartile	65	64	63	62	61
Owners 3 <sup>rd</sup> Quartile	70	71	72	73	74
Owners 4 <sup>th</sup> Quartile	85	84	83	82	81
Residual Vacant	99				

All addresses in the Unit frame fall into one of these cells. When the addresses are placed in order, those whose rent is in the lowest quartile and have a small number of occupants are at one end of the list, and those whose property values are in the highest quartile with a small number of occupants are at the other end of the list. The stratification code is a surrogate for sorting by expenditures. To draw a systematic sample, the Unit frame addresses are sorted by PSU, urban/rural classification, FIPS State code, FIPS County code, the stratification variable described above, Census Tract code, Census Block code, Basic Street Address, and Unit Sort Order code.

Addresses in the Area frame are sorted in a similar way, but using different variables. A Combined Block Stratification code is calculated using median household size and the proportion of owner-occupied housing units. Then addresses in the Area frame are sorted by PSU, an urban/rural variable, Combined Block Housing Flag, FIPS State code, FIPS County code, Combined Block Stratification code, Census Tract, and a Combined Block code.

The GQ and Permit frames do not have a stratification code, but they have a within-PSU sort. The sort variables in the GQ frame are: PSU, FIPS State code, FIPS County code, Census Tract, Combined Block code, and a Within-Combined-Block code.

For more information on sampling within PSUs for the CE Surveys, please refer to the 2008 CE Anthology article, “Selecting a Sample of Households for the Consumer Expenditure Survey” by Susan King and Sylvia Johnson-Herring (Attachment O).

### Estimation

The estimation procedure for both the CED and CEQ follow well-established statistical principles. The final weight for each sample CU is the product of the inverse of its probability of selection; a weight adjustment to account for noninterviews; and a calibration adjustment that post-stratifies the weights to account for population undercoverage.

For additional information on the sample design and estimation methodology used in the CE surveys, please refer to “Chapter 16, Consumer Expenditures and Income” in the *BLS Handbook*

*of Methods* (Attachment Q). Another source of information is Kenneth V. Dalton's memo to Chester E. Bowie, "Specifications for the Selection of CE/CPI Samples in PSUs Based on the 2000 Census," June 28, 2002 (Attachment R); and Alan R. Tupek's memo to Kenneth V. Dalton, "Calculations of Within-PSU Sampling Intervals for the Census 2000-Based Redesign of the Consumer Expenditure Surveys and the CPI Permit New Construction Housing Sample," November 11, 2002 (Attachment S).

### 3. **Methods to Maximize Response Rates**

In the CE Surveys, keeping the noninterview rate at a low level requires special efforts, particularly from the Census Bureau Field staff. For each refusal case, the regional office sends a special letter to the address and assigns the case for follow-up by the program supervisor, supervisory field representative, or senior interviewer, taking into account time and cost considerations.

To adjust for those noninterviews that the field staff cannot convert to interviews, the sample design provides for a noninterview adjustment in the estimation procedure. The computer processing employs special techniques in the CEQ to reference data provided in the previous interview, keeping recall problems and interview time to a minimum.

### 4. **Testing Plans**

Subject to resource availability, CE plans to conduct the following studies (prior to the expiration of the clearance). Ideally these studies will utilize non-production sample, but funding may necessitate the use of production sample for some tests. A Non-Substantive Change Request (NCR) will be submitted for all of the proposed studies should funding and resources become available.

#### CE Interview Survey

- 1) Telephone Friendly Questions - A field test of questions that are more neutral between personal visit and telephone interviewing. Current questions in the interview were designed for a personal visit setting, relying on long lists from the Information Booklet and do not convey well in telephone interviewing.
- 2) Promotional Materials Test - A test of providing promotional materials to respondents in order to encourage participation in subsequent waves. A maximum of three items will be tested in several regions to determine the effectiveness of providing the promotional materials to respondents. The items that have been considered for this test are: ceramic, acrylic or plastic mugs, magnets, and reusable canvas bags all with the Census and BLS logos.
- 3) Expense Worksheet - This study is designed to determine whether it is feasible to ask respondents to record expenditures between interviews on a worksheet and to determine if recording expenditures between interviews results in better data quality and potentially less respondent burden as preparation may make the interview shorter.
- 4) Information Booklet for Telephone Respondents Test – This test will follow up on previous research conducted to determine if providing an Information Booklet to telephone respondents improved data quality. Initial research showed that it was not feasible to provide an Information Booklet to telephone respondents using the prior test procedures. Also, there was

some evidence that use of the Information Booklet during a telephone interview may have affected the level of expenditures reported. This test would allow CE to test another set of field procedures and better determine the affect of the booklet on expenditure reporting levels.

## CE Diary Survey

- 1) Individual Diary Feasibility Study – This study, as described earlier, will test the feasibility and impact of using both individual diaries and multiple modes to collect CE Diary expenditures. All CU members in the research sample will be offered, sequentially, a mobile-optimized web option, then a web survey, and then a two-week paper diary for recording expenditures. The purpose of this test is to evaluate the operational issues related to implementing a mobile-optimized web option for Diary data collection versus web mode with an individual log-in versus a two-week paper diary; to evaluate if high-level expenditure reporting rates and data quality differ by research group and the control group; to determine differences in acceptance and completion rates by research group and control group; and lastly, to ascertain web respondents’ and mobile-optimized web option respondents’ specific data entry patterns.
- 2) Expenditure App Prototype Study – This study includes the development and a small scale test of CE specific Smartphone applications (apps). A contract will be issued to a private vendor to develop iOS and Android apps designed to collect CE specific data elements. The apps will be branded as Consumer Expenditure Survey applications. Internal studies of existing apps were conducted by BLS and none easily collected all required CE data elements. The apps will be evaluated as both primary and secondary sources of CE data.
- 3) TPOPS Test - A test to combine the CE Diary Survey and CPI’s Point of Purchase Survey (TPOPS). This test might be undertaken to determine whether data collected on point of purchase can be combined with expenditure data in order to provide CPI with lower levels of nonresponse error for point of purchase data.

## 5. Statistical Contacts

The Census Bureau will collect the data. Within the Census Bureau, you may consult the following individuals and their area of expertise for further information.

Sample Design:	Stephen Ash	(301) 763-4294
Data Collection:	Howard McGowan	(301) 763-5342