

## Supporting Statement

### B. Collection of Information Employing Statistical Methods

#### 1. Sampling Method

##### Consumer Units

There are approximately 120 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	2,638,364
A103	3,026,287
A104	1,090,977
A109	3,414,216
A110	3,290,790
A111	2,840,143
A207	3,910,398
A208	2,326,271
A209	1,142,158
A210	1,255,913
A211	1,337,074
A312	2,143,530

A313	1,088,433
A316	2,249,314
A318	2,010,347
A319	1,791,131
A320	1,652,640
A321	1,001,219
A419	5,271,911
A420	1,708,762
A422	3,001,133
A423	1,515,522
A424	1,199,638
A425	1,080,483
A426	416,102
A427	151,786

A429	1,386,391
A433	1,249,024
X102	1,952,217
X104	1,134,768
X108	1,502,043
X210	1,516,051
X212	1,014,829
X218	1,156,950
X220	1,170,958
X222	1,285,116
X224	1,836,414
X226	1,037,673
X228	805,277
X232	745,161

<sup>11</sup> The number of CUs comes from dividing the Census Bureau's 2009 estimate of the number of people in the civilian non-institutional population (300 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 120 million CUs by each stratum's proportion of the nationwide population in the 2000 Census.

X336	1,688,923
X338	1,174,014
X340	1,209,325
X342	1,572,091
X344	1,230,983
X346	1,114,963
X350	945,564
X352	1,207,497
X354	1,738,799
X356	1,227,421
X358	1,721,728
X362	1,572,044
X364	1,753,919
X366	889,438
X368	1,231,919
X472	1,751,462
X474	1,976,543
X482	1,479,791
X484	1,293,500
Y102	580,120
Y104	674,484
Y206	850,575
Y208	1,011,573
Y210	843,278
Y212	1,011,291
Y314	837,912
Y316	968,028
Y318	802,863
Y320	845,327
Y322	999,410
Y324	800,090
Y426	544,372
Y428	497,626
Y430	613,404
Y432	593,330
Z102	381,318
Z104	593,145
Z206	875,887
Z208	644,053
Z210	791,953
Z212	940,569
Z314	797,111
Z316	756,399
Z318	801,517
Z320	1,052,338
Z322	972,994
Z324	578,927

Z426	279,795
Z428	243,107
Z430	335,915
Z432	353,952
<b>Total</b>	<b>120,000,000</b>

## Response Rates

The following table shows expected 2011 sample sizes for the Quarterly Interview Survey (CEQ) and the Diary Survey (CED) based on the target number of completed interviews and 2009 response rates.

Category	Quarterly (quarter) (1)	Diary (annual)
Total Sample Size	14,725	12,075
Total Type B and C Noninterviews (vacant, demolished, etc.): Number	2,950	2,425
Percent of Total Sample	20.0	20.0
Total Eligible Units	11,775	9,650
Total Type A Noninterviews Number	2,950	2,600
Percent of Total Eligible	25.0	27.0
Total Completed Interviews Number	8,825	7,050
Percent of Total Eligible	75.0	73.0

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

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 V).

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).

## 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:

<b>PSU “size class”</b>	<b>Number of PSUs</b>	<b>Description</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.

The CEQ and CED use the same sample designs. They collect data in the same PSUs, and they share a single sample of households drawn from the survey’s sampling frames. Within each PSU a single systematic sample of households is drawn from each sampling frame for the two surveys, with the even-numbered “hit strings” assigned to the CEQ and the odd-numbered “hit strings” assigned to the CED. Thus both surveys try to collect data from the same number of households. The difference in sample sizes shown in CE publications is due to the fact that each household in the CEQ is asked to participate in four nonbounding interviews, while each household in the CED is asked to complete two weekly diaries. That is why the sample sizes shown in most CE publication are approximately in the ratio of four-to-two (or equivalently two-to-one).

### 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 extreme, and those whose property values are in the highest quartile with a small number of occupants are at the other extreme. 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**

As part of CE’s goal of continuous improvement, the following methodological studies will be tested (prior to the expiration of the clearance) using the production sample, pending funding and resource availability. Should funding and resources become available, a Non-Substantive Change Request (NCR) will be submitted for all of the proposed studies.

#### CEQ Interview

- 1) \$40 CEQ Incentive – Continue to study the results of the prior CEQ Incentives test and determine if additional testing is needed. The results suggest that offering a \$40 incentive improves response rates, data quality, respondent cooperation, sample composition, and income reporting.
- 2) Bounding Effect Study - A test of bounding effects, by removing sections of the first wave questionnaire in an effort to decrease respondent burden;
- 3) Telephone Friendly Questions - A field test of questions that are mode 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.

- 4) 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.

#### CED Diary

- 1) Web Diary Test - Test of a web diary designed to decrease respondent burden by offering respondents an electronic way to report expenditures;
- 2) Individual Diaries Test – An individual diary test designed to increase reported expenditures by collecting data from all consumer unit members. This test will most likely be combined with the web diary test.

In addition, non-production samples will be used to test the following CEQ Interview topics:

- 1) TPOPS Test - A test to combine the CE Interview 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.
- 2) An Expense Worksheet Test. 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.
- 3) 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.

## 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