B. Collection of Information Employing Statistical Methods

1. **Sampling Method**

   Consumer Units
   There are approximately 124 million consumer units (CUs) in the Consumer Expenditure (CE) Survey's universe.[1] A CU is the unit from which CE seeks 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. For unrelated persons it 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 Memorandum from Jay Ryan on PSUs for the Consumer Expenditure Survey's 2010 Census-Based Sample Design (Attachment T).

   The following table shows estimated numbers of CUs in all 91 strata from which PSUs were selected.[2] Please see the section 2 below entitled "Primary Sampling Units (PSUs)" for more information.

Estimated Number of CUs in CE's 91 Strata

Stratum Code	Estimated Number of CUs in the Stratum
S11A	1,828,360
S12A	7,858,766
S12B	2,395,832
S23A	3,799,819
S23B	1,725,482
S24A	1,344,986
S24B	1,119,611

S35A	2,263,653
S35B	2,234,898
S35C	2,123,283
S35D	1,117,821
S35E	1,088,601
S37A	2,580,930
S37B	2,377,788
S48A	1,683,969
S48B	1,021,527
S49A	5,152,385
S49B	1,741,202

S49C	1,696,807
S49D	1,381,514
S49E	1,243,156
S49F	546,331
S49G	210,112
N11B	2,010,453
N11C	1,700,452
N12C	1,632,959
N12D	1,398,931
N12E	1,576,507
N12F	1,430,722

---

[1] The number of CUs comes from dividing the Census Bureau's 2012 estimate of the number of people in the civilian non-institutional population (310 million) by the average number of people per CU (2.5).

[2] The number of CUs per stratum comes from allocating the nationwide total of 124 million CUs by each stratum's proportion of the nationwide population in the 2010 Census.

N23C	1,363,860		N49K	1,752,562
N23D	1,308,476		R11D	262,158
N23E	1,509,512		R12G	331,690
N23F	1,307,890		R23K	644,884
N23G	1,576,106		R23L	542,780
N23H	1,570,832		R24G	738,217
N23I	1,504,137		R24H	621,636
N23J	1,376,516		R35R	619,716
N24C	1,194,440		R35S	744,494
N24D	1,141,728		R36G	629,641
N24E	1,320,671		R36H	565,076
N24F	1,183,952		R37K	528,297
N35F	1,218,992		R37L	637,760
N35G	1,061,472		R48G	193,447
N35H	1,216,063		R48H	160,385
N35I	1,023,813		R48I	179,683
N35J	1,242,837		R49L	286,919
N35K	1,059,119		**Total**	**124,000,000**
N35L	1,241,485			
N35M	1,031,672			
N35N	1,169,991			
N35O	1,098,976			
N35P	1,245,281			
N35Q	1,029,405			
N36A	1,015,961			
N36B	997,479			
N36C	1,052,497			
N36D	1,125,113			
N36E	1,024,308			
N36F	962,821			
N37C	978,397			
N37D	1,129,751			
N37E	1,021,578			
N37F	981,908			
N37G	1,036,610			
N37H	1,106,926			
N37I	1,052,659			
N37J	1,145,412			
N48C	1,296,430			
N48D	1,495,761			
N48E	1,542,522			
N48F	1,287,915			
N49H	2,091,811			
N49I	2,073,860			
N49J	1,856,850			

Response Rates

The following table shows expected annual sample sizes for the Quarterly Interview Survey (CEQ) and the Diary Survey (CED) for 2015 and after. A new sample design is starting for both surveys in 2015. The CEQ has two columns, "2015" and "After 2015." The first column is for the transition period from the old sample design to the new sample design, and the second column is for the new sample design alone. The CED has only one column and it is for the new sample design alone. Unlike the CEQ, the CED does not require a transition period.

The "Type B/C rate" is the percent of sample addresses that are not occupied housing units – they are nonexistent, nonresidential, vacant, demolished, etc. The rest are occupied housing units, and the "Type A rate" is the percent of those occupied housing units that did not prticipate in the survey.

The response rates shown below are the CEQ's and CED's actual response rates over the past five years (2008-2012) minus 5 percentage points. Response rates have been decreasing over time, so the 5-year historical response rates are reduced by 5 percentage points to account for this downward trend.

Starting in 2015 the CEQ and CED will draw their samples of addresses from a new sampling frame called the Master Address File (MAF), which is a list of addresses from the 2010 census plus biannual updates from the U.S. Postal Service's Delivery Sequence File. The CEQ and CED do not have any experience with the MAF, but the ACS does have some experience with the MAF, and the Type B/C rate of 13% comes from ACS's experience. The Type B/C rate of 16% for the CEQ's 2015 transition period is a weighted average of the Type B/C rates from the new and old sampling frames (13% and 20%).

The sample sizes shown below are the annual number of quarterly interviews for CEQ, and the annual number of bi-weekly diaries for CED.

| | Quarterly Interview | | Diary |
Category	2015	After 2015	2015 and After
Total Sample Size	50,400	48,000	12,000
Type B and C Noninterviews (vacant, demolished, etc.)			
Number	8,064	6,240	1,560
Percent of Total Sample	16.0	13.0	13.0
Eligible Units			
Number	42,336	41,760	10,440
Percent of Total Sample	84.0	87.0	87.0
Type A Noninterviews			
Number	14,394	14,198	3,550

Percent of Eligible Units	34.0	34.0	34.0
Completed Interviews			
Number	27,942	27,562	6,890
Percent of Eligible Units (Response Rate)	66.0	66.0	66.0

The sample size for 2012 and 2013 for Interview were approximately 61,280 annually and for Diary, the sample size in 2012 and 2013 were approximately 12,700.

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

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 an OMB directive. Results from four individual studies were synthesized, and they concluded that no bias was generated in spite of that fact that CE's data are not "missing completely at random." 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, see "Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey" (Attachment R).

2. **Collection Methods**
   Under contract with BLS, field representatives from the U.S. Census Bureau personally visit the households in the Diary and Interview surveys' samples to collect the data. Prior to a household visit, respondents in both the Diary and Interview surveys' are sent an advanced letter. Each household in the Diary survey is asked to record all of the expenditures it makes during a 2-week period. Field representatives visit each household in the sample three times. On the first visit, the field representatives introduce themselves, explains the survey, and leaves a diary in which the household members are asked to record all their expenditures for a 1-week period. On the second visit, the field representatives pick up the first week's diary, ask whether there are any questions, and leave another diary for the second week. On the third visit, the field representatives pick up the second week's diary and thank the household for participating in the survey. After participating in the survey for 2 weeks, the household is dropped from the survey, and it is replaced by another household. Each household in the Interview survey is interviewed every 3 months for 4 consecutive quarters. Trained field representatives ask household members about their expenditures over the previous 3 months. Responses are entered into a laptop computer. The households in the Interview survey are on a rotating schedule, with approximately one-fourth of the households in the sample being new to the survey each quarter. Households in both the Interview and Diary survey are sent a Thank You letter, as well as a certificate of appreciation, after completion of the Interview or second week of the Diary.

   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 the PSUs selected for the sample 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 three categories:

PSU "size class"	Number of PSUs	Description
S	23	Large Metropolitan Core Based Statistical Areas (self-representing PSUs)
N	52	Small Metropolitan Core Based Statistical Areas and Micropolitan Core Based Statistical Areas (non-self-representing PSUs)
R	16	Non- Core Based Statistical Areas 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 N and R PSUs) used a 4-variable model whose independent variables were latitude, longitude, median household income, and median household property value.

Sampling Within PSUs

Two non-overlapping sampling frames are used to select CUs for the CEQ and CED. They are the Unit and Group Quarters frames. The Unit frame is the Census Bureau's new "Master Address File" (MAF), and it covers more than 99% of the sample. It is a list of every residential address identified in the 2010 decennial census plus biannual updates from the U.S. Postal Service's Delivery Sequence File. The Group Quarters frame covers less than 1% of the sample and it consists of boarding houses, hotel rooms, and institutions that were found in the decennial census but are not counted as housing units. It also consists of new group quarters found via the MAF's biannual updates, field operations, and the new college housing survey.

A "systematic sample" of households is selected from the two frames in each PSU. 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. Then after the initial household is selected every k-th household down the list is selected where "k" is the PSU's sampling interval.

The two frames have 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 1st Quartile	12	10	11	13	14
Renters 2nd Quartile	23	25	24	22	21
Renters 3rd Quartile	32	30	31	33	34
Renters 4th Quartile	43	45	44	42	41
Owners 1st Quartile	52	50	51	53	54
Owners 2nd Quartile	63	65	64	62	61
Owners 3rd Quartile	72	70	71	73	74
Owners 4th Quartile	83	85	84	82	81
Other	99				

All addresses in the Unit frame fall into one of these cells. When the addresses are sorted by this stratification code, renters whose rent is in the lowest quartile are at one end of the list, and owners whose property values are in the highest quartile 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, FIPS State code, FIPS County code, the stratification variable described above, Census Tract code, Census Block code, U.S. Postal Service Zip code, Basic Street Address, and MAFID code.

The Group Quarters frame does not have a stratification code, but it has a within-PSU sort. The sort variables in this frame are: PSU, FIPS State code, FIPS County code, Census Tract code, Census Block code, and Census "half-measure" code. (This "half-measure" code identifies two of every four units in a cluster.)

CE interviews consumers who live in non-institutional group quarters, and does not interview consumers who live in institutional group quarters.

Institutional group quarters are primarily correctional facilities or nursing homes, whose residents are formally classified as "inmates or patients." Typically, these people stay involuntarily and cannot come and go without permission and are generally under the supervision of trained staff.

Non-institutional GQs house people who stay voluntarily and are allowed to come and go without receiving permission. Many non-institutional GQs are college housing; dormitories; and fraternity and sorority housing, both on and off campus. In addition, non-institutional GQs include: hotels and motels that are used entirely or partially for persons without a usual home; shelters for the homeless with sleeping facilities.

Military quarters, with the exception of military disciplinary barracks [stockades and jails], are also categorized as non-institutional group quarters. However, only the non-institutional civilian population as opposed to military personnel is eligible to participate in the CE survey. Therefore, military non-institutional group quarters can be listed if and only if the GQ includes non-institutional, non-military units.

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 its base weight (which is the inverse of the CU's probability of selection); a weight adjustment to account for noninterviews; and a calibration adjustment that post-stratifies the weights to account for population undercoverage. A typical base weight for a CU in the CEQ is approximately 10,000, which means it represents 10,000 CUs – itself plus 9,999 other CUs that were not selected for the survey. A typical final weight for a CU in the CEQ is approximately 18,000, which means it represents 18,000 CUs – itself plus 17,999 other CUs that were not selected for the survey and/or did not participate in the survey.

For additional information on the sample design and estimation methodology used in the CE surveys, refer to "Chapter 16, Consumer Expenditures and Income" in the *BLS Handbook of Methods* (Attachment S); Jay Ryan's memo to Richard Schwartz, "PSUs for the Consumer Expenditure Survey's 2010 Census-Based Sample Design," December 18, 2012 (Attachment T); and Ruth Ann Killion's memo to Jay Ryan, "Consumer Expenditure Surveys Sample Allocation for the 2010 Census-Based Sample Design," February 11, 2014 (Attachment U).

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, to keep 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.

Test	Survey	Description
Proof-of-Concept Main Data Collection	Interview and Diary	The Proof-of-Concept test will examine the approved proposal for the redesign for the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CE) on a small sample, by assessing potential operational, logistical, and selected data quality issues.
Outlet Questions Testing	Interview	The testing of outlet questions in production sample will assess the burden and data quality trade-offs of meeting Consumer Price Index (CPI) Telephone Point of Purchase (TPOPS) requirements in the CE survey.
Incentives Field Test	Interview and Diary	The Incentive Field Test will evaluate differential incentive amounts and structure based on other examples from the statistical agency community and previous CE tests. The test will include a non-monetary incentive component comprised of a package of information provided to respondents.
Large Scale Feasibility Pilot Test	Interview and Diary	The Large Scale Feasibility Test will evaluate the updated redesign for the CE on a large sample, by assessing potential operational, logistical, and data quality issues. The pilot test will be undertaken to evaluate the protocol before fielding of main data collection.

5. **Statistical Contacts**

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

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