Review of the 2010 Sample Redesign of the Consumer Expenditure Survey October 2015

Danielle Neiman[1], Susan King[2], David Swanson[2], Stephen Ash[1],
Jacob Enriquez[1], Joshua Rosenbaum[1]
[1]U.S. Census Bureau, 4600 Silver Hill Rd, Washington DC 20233
[2]U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington DC 20212

Abstract
The Consumer Expenditure Survey (CE) is a nationwide household survey conducted jointly by the U.S. Bureau of Labor Statistics and the U.S. Census Bureau to investigate how Americans spend their money. Every ten years the survey updates its sample of geographic areas around the country as well as its sample of households in those geographic areas based on the latest decennial census to ensure the sample accurately reflects shifts in the American population. This paper describes CE's latest sample design that will be used over the next ten years (2015–2024), including research that went into its decisions. Topics include the coordination of CE's household sample with other household surveys conducted by the Census Bureau, and a new annual sampling methodology used by all Census Bureau household surveys.

Key Words: Sample-design stratification, sample selection, sample allocation, first-stage sample design, second-stage sample design, Consumer Expenditure Survey

1. Introduction

The Consumer Expenditure Survey (CE) is a nationwide household survey which collects data on the expenditures made by American households. After every decennial census, CE redesigns its survey to reflect population changes, and to improve both coverage and sample selection procedures. This paper explains the new design implemented in 2015 which uses the 2010 Decennial Census (Design 2010) and compares it to the design implemented in 2005 after the 2000 Decennial Census (Design 2000).

1.1 Survey Description

The CE Survey consists of two independent surveys: the CE Interview Survey and the CE Diary Survey. The CE Interview Survey collects detailed expenditure data on large expenditures such as property, automobiles and major appliances; and on recurring expenditures such as rent, utilities, and insurance premiums. Each household is interviewed every three months for four consecutive quarters by a field representative from the U.S. Census Bureau. Each interview takes approximately an hour. Conversely, the CE Diary Survey collects detailed expenditure data on small, frequently purchased items such as food and apparel. A household completes two one-week diaries requiring three visits from the field representative. Both surveys share the same sample design.

CE data is used in a variety of ways. The Consumer Price Index (CPI) is the primary customer of the CE Survey and uses consumer expenditure data to select new "market baskets" of goods and services for the index, to determine the relative importance of its components, and to derive cost weights for the baskets. CE also uses consumer

expenditure data to calculate poverty thresholds for the Supplemental Poverty Measure, which is an additional measure and not the official poverty measure. The Internal Revenue Service uses consumer expenditure data to calculate alternate sales tax standard deductions. The Department of Defense uses consumer expenditure data to determine cost-of-living allowances for military personnel living off military bases. Also, market researchers find consumer expenditure data valuable in analyzing the demand for various groups of goods and services.

1.2 Overview of the Sample Selection Process

CE's universe of interest is the U.S. civilian non-institutional population, which includes people living in houses, condominiums, apartments, and group quarters such as college dormitories. However, military personnel living on base, nursing home residents, and prison inmates are excluded. The civilian non-institutional population represents more than 98 percent of the population of the United States. The unit of interest is the consumer unit, a group of people who pool their incomes to make joint expenditure decisions. Consumer units include families; groups of unrelated people who live together and pool their incomes to make joint expenditure decisions; and single persons who live alone or with other individuals but who are financially independent of the other individuals. There can be multiple consumer units in a household, but generally a consumer unit and a household are equivalent.

CE uses a two-stage sample design to select a sample of households from the civilian non-institutional population. In the first-stage, single counties or groups of adjacent counties are assigned to Primary Sampling Units (PSUs). There are two types of PSUs: urban and rural. Every county in the United States is assigned to an urban or a rural PSU, but only a subset of the PSUs is selected for sampling. As mentioned above, a primary customer of CE is the Consumer Price Index (CPI) and the two surveys worked together on selecting a common set of urban PSUs in Design 2010, allowing CPI to collect prices in the same areas that CE collects expenditure data. CE also selects a sample of rural PSUs to collect household expenditure data, but CPI does not collect prices in rural areas.

In the second-stage of the sample design, addresses are selected by systematic sampling within each PSU. The second-stage design is a joint effort by the Bureau of Labor Statistics and the Census Bureau and the selected households are interviewed by a representative of the Census Bureau. The addresses are selected in conjunction with other household surveys including the Current Population Survey (CPS), Survey of Income and Program Participation (SIPP), National Crime Victimization Survey (NCVS), and American Housing Survey (AHS) which make up the Demographic Household Surveys of the Census Bureau. The Demographic Household Surveys share the same sampling frames and the same sampling systems to minimize overlap between the surveys and to reduce the probability that a household is asked to participate in multiple surveys during the lifetime of the design.

1.3 Changes in the Survey

Previously, each household in the Interview Survey was interviewed every three months for five consecutive quarters. The first interview was used only for "bounding" purposes to address a common problem in which survey respondents tend to report expenditures more recent than actually occurred. The bounding interview was never used in calculating expenditure estimates and was dropped to reduce respondent burden and the survey's cost (Ryan 2013). In Design 2010, there are only four consecutive interviews.

Another change is the source of demographic variables used in the creation of the new sample design. In Design 2010, both stages of the sample design switched to using demographic variables found in the American Community Survey (ACS), a continual monthly survey, which makes frequent demographic updates possible. Previously, the demographic variables came from the long form of the decennial Census, which is no longer conducted.

2. First Stage Sample Design: Defining and Selecting a Sample of PSUs

There are three major tasks in the first stage of a multi-stage stratified sample design: defining PSUs, stratifying PSUs, and selecting PSUs (Murphy 2008).

2.1 Defining PSUs

The U.S. Office of Management and Budget (OMB) assigns counties surrounding an urban core to geographic entities called Core Based Statistical Areas (CBSAs). The assignment is based on each county's degree of economic and social integration to the urban core as measured by commuting patterns. There are two types of urban CBSAs: metropolitan and micropolitan. A metropolitan CBSA has an urban core with more than 50,000 people and a micropolitan CBSA has an urban core of between 10,000 and 50,000 people. CBSAs form the urban PSUs in the CE Survey and may cross state borders.

Counties which are not part of a metropolitan or micropolitan CBSA are rural and are sampled by CE. Since OMB does not group rural counties into small clusters of adjacent counties, CE defines its own PSUs. CE requires a rural PSU to be within a state border, to consist of adjacent rural counties, have a land area less than 3,000 miles and have a minimum population of 7,500 people. The last two constraints are guidelines used by the Census Bureau for establishing the maximum workload for a single field representative (Murphy 2008). Prior to Design 2010, there was no formal procedure for assigning adjacent rural counties to a PSU, so an algorithm was developed for Design 2010 using an adjacency matrix and zero-one integer linear programming (King 2012).

2.2 Self-Representing and Non-Self Representing PSUs

All 3,143 counties in the United States are assigned to a PSU and each PSU is assigned to a stratum based on its size-class. Then one PSU is selected to represent all of the PSUs in the stratum with probability proportional to size. Very large metropolitan PSUs are assigned to their own stratum and are selected with probability of one. Consequently, these PSUs are referred to as self-representing. In Design 2010, self-representing PSUs have populations greater than 2.5 million people, whereas in Design 2000, the population cut-off was 2.7 million. The remaining PSUs are non-self-representing. In Design 2010, the non-self-representing metropolitan and micropolitan PSUs are stratified together. The rural PSUs have their own stratum in both designs.

In Design 2010, the self-representing PSUs are called "S" PSUs; the non-self-representing metropolitan and micropolitan PSUs are called "N" PSUs; and the rural PSUs are called "R" PSUs. In Design 2000, the self-representing PSUs were called "A" PSUs; the non-self-representing metropolitan PSUs were called "X" PSUs; the micropolitan PSUs were called "Y" PSUs; and rural PSUs were called "Z" PSUs. Thus, the number of size-classes was reduced from four to three in the new sample design.

For stratification, Alaska and Hawaii are separated from the continental United States because they have homogeneous markets with unique pricing behaviors and weak

correlation with price changes of the other non-self-representing PSUs in the western United States. For this reason, in the earlier designs, both Anchorage, AK and Honolulu, HI were self-representing PSUs even though their populations were below the cut-off. In the new design, the four CBSAs in Alaska were grouped into a state stratum and Anchorage was selected to represent the state stratum. Likewise, the four CBSAs in Hawaii were grouped into a state stratum and Honolulu was selected to represent the stratum.

The total number of self-representing and non-self-representing PSUs in the sample are determined by budgets and other factors including sampling variance and bias. Based on these criteria, it was decided that both CE and CPI would have 75 urban PSUs and CE would have 16 rural PSUs, which is the same as CE's Design 2000. There are 23 self-representing PSUs including Anchorage and Honolulu and 52 non-self-representing PSUs in the sample. The 52 stratification clusters for those PSUs are divided among the nine Census Divisions.

2.3 Stratifying Non-Self-Representing PSUs

The non-self-representing PSUs are stratified by size-class and geographic division. There are two size-classes which are the "N" and "R" categories mentioned above, and nine geographic divisions. The Census Bureau divides the United States into four geographic regions (Northeast, Midwest, South, and West), and each region has two divisions except the South which has three divisions, which makes a total of nine divisions. Previously CE and CPI stratified by region, but stratifying by division allows the CPI to increase the number of inflation rates it publishes. Then, after the non-self-representing PSUs are stratified within their size-class and geographic division, one PSU per stratum is randomly selected to represent the stratum.

The primary objective of PSU stratification is to minimize the between-PSU component of sampling variance (Murphy 2008). In other words, the PSUs within each stratification cluster should be as homogenous as possible with respect to the survey variable, expenditures, but there should be variability between the stratification clusters. Also, within each division, each stratification cluster should have approximately the same population to minimize variance. This is a constrained clustering problem and is solved using heuristic algorithms. Traditional clustering algorithms find homogenous stratification PSUs, but do not balance the population. In the previous design, the Friedman-Rubin hill climbing algorithm was used to assign PSUs to stratification clusters, but in Design 2010 a new heuristic stratification algorithm was developed which uses k-means clustering and zero-one integer linear programming (King et al., 2011). In Design 2010, four clustering variables were used: median household income, median household property value, latitude and longitude. Median household income and median property value correlate with expenditures and are calculated for each PSU from five-year ACS estimates.

2.4 Selecting Non-Self-Representing PSUs

After the non-self-representing PSUs are assigned to stratification clusters, 16 rural PSUs are selected with probability proportional to size to represent their strata. However, the 52 non-self-representing metropolitan and micropolitan PSUs are stratified together and selected using maximum overlap and controlled selection.

Since there are significant costs both financial and in loss of expertise when opening and closing field offices, it is desirable to retain as many of the current PSUs as possible in

the new sample (Ernst et al., 2007 and Johnson et al., 2012). Maximum overlap procedures attempt to retain as many of the PSUs from the old sample design as possible and are done in a way that preserves the unconditional selection probabilities in the new design. In Design 2010 it was conducted at the stratum level. All overlap methods create and use conditional probabilities based on their overlap rules. Only non-self-representing metropolitan PSUs are overlapped in Design 2010. All of the PSUs in the stratum are used in the calculation of the conditional probabilities of selection. In Design 2000, the Perkins (1970) method of maximum overlap, a heuristic procedure, was used, whereas in Design 2010, the Ernst (1986) method, which uses linear programming, was used. The Ernst method determines the set of conditional probabilities that maximize the expected unconditional number of PSUs that will be re-selected. The two procedures have different assumptions and the overlap is larger in the Ernst method, lowering the cost of the new design.

The actual sample of PSUs is selected using controlled selection, and it is based on the PSUs' conditional probabilities that were derived from the overlap maximization process described above. In each Census Region, there are several strata and one PSU is selected from each stratum. Certain combinations of PSUs or patterns are preferred because they lower the sample variance or more evenly distribute the sample according to constraints such as the number of PSUs per state, or the percentage of micropolitan and metropolitan PSUs in the region. Thus, controlled selection controls for interaction between PSUs across strata by increasing the probability of selecting a preferred pattern. Mathematical optimization techniques are often used in control selection. In Design 2010, the non-self-representing metropolitan and micropolitan PSUs are in the same stratum. Since CPI found a difference in price change behavior in metropolitan and micropolitan areas, the number of metropolitan and micropolitan PSUs are controlled. In Design 2000, controlled selection was performed at the Census Region level and controls were on the number of overlap PSUs and PSUs per state.

Although, conditional probabilities are used in the overlap maximization and controlled selection process to select the 52 non-self-representing PSUs for the sample, the unconditional probability of selection, the selected PSUs population divided by its stratum population, is used in weighting.

3. Second-Stage Sample Design: Selecting a Sample of Households

Once a sample of PSUs is selected, the next stage of the sample design is selecting a representative sample of households within the PSUs. This involves several sub-steps, which include: determining the survey's total nationwide sample size based on the survey's total available budget, allocating the sample to all of the individual PSUs, and selecting a systematic sample of addresses. The goal of this process is to select a sample which minimizes the variance of CE's most important statistic, the average annualized expenditure per household nationwide on all items.

There are many second-stage changes to Design 2010. In prior designs, the civilian non-institutional population was represented by four frames and those frames were shared by the Demographic Household Surveys[1] of the Census Bureau. In Design 2010, the

[1] The Demographic Household Surveys of the U.S. Census Bureau include the Current Population Survey, Survey of Income and Program Participation, American Housing Survey, and the National Crime Victimization Survey.

Demographic Household Surveys made the decision to move towards a two frame sample design which incorporates annual sampling and moves away from the once-a-decade sampling of Design 2000. Another change from the last design was the discontinuation of the decennial census long-form which caused a change in the variables used to stratify households in the systematic sample. The new variables are from the ACS and this new process allows more up-to-date information about the U.S. population to be included in the sample selection process annually. Also, the optimization program used to select the sample size for each PSU was modified.

3.1 Sample Allocation and Sample Size

The first sub-step of selecting a sample of addresses within each PSU is determining the survey's nationwide sample size and allocating it to the sample PSUs. CE's budget allows 12,000 addresses to be selected per year for the Interview Survey and 12,000 addresses per year for the Diary Survey.

The objective of the allocation process is to allocate the 12,000 addresses to the PSUs in a way that minimizes CE's nationwide variance. It uses a two-step population-based technique: stratify the 91 sample PSUs into 41 "index areas" defined by CPI, allocate the nationwide sample of 12,000 addresses directly proportional to the population represented by each of the CPI index areas, and then sub-allocate the sample to individual PSUs in the index areas. The 41 index areas consist of the 23 self-representing PSUs plus the 18 non-self-representing division size-classes (9 Census divisions x 2 size-classes). This model was first used in Design 2000, and recent research by BLS and Census confirmed that this method is still the simplest and most effective way of producing expenditure estimates with small variances at the nationwide level (Swanson et al., 2011 and 2012).

The allocation is accomplished by solving the following nonlinear optimization problem:

Given the values of p_i and r_i for every index area *i*, find the values of n_i that

$$\text{Minimize} \quad \sum_{i=1}^{41} \left(\frac{n_i r_i}{NR} - \frac{p_i}{p} \right)^2$$

$$\text{Subject to:} \quad \sum_{i=1}^{41} n_i = 12{,}000$$

$$n_i r_i \geq 80, \text{ for } i = 1 \; to \; 32$$

$$n_i r_i \geq 40, \text{ for } i = 33 \; to \; 41$$

where
p_i = population of the *i*-th index area;
r_i = participation rate (eligibility rate times the response rate) of the *i*-th index area;
n_i = number of addresses allocated to *i*-th index area;
$p = \sum_{i=1}^{41} p_i$ is the population of the United States;
$n_i r_i$ = expected number of interviewed households in the i-th index area;
$NR = \sum_{i \in USA} n_i r_i$ is the expected number of interviewed households nationwide.

As mentioned above, CE's budget allows 12,000 addresses to be selected per year for the Interview Survey and 12,000 addresses per year for the Diary survey. The objective is to

allocate the 12,000 addresses in a way that minimizes CE's nationwide variance. The objective function shown above minimizes the sum of squared differences between each index area's share of the national population and its share of the addresses, which is a good approximation to minimizing the nationwide variance. The total U.S. population, p, is known as well as the population of each index area, p_i. The expected number of interviewed households is $n_i r_i$, where n_i is the number of addresses and is the decision variable to be determined in the optimization model and r_i is the participation rate for index area i. The total number of interviewed households is *NR*. The first constraint is linear and restricts the number of addresses to 12,000. The lower bound constraints require at least 80 addresses in each of the 32 urban index areas (i = *1 to 32*) and 40 addresses in each of the 9 rural index areas (i = *33 to 41*).

The participation rate is the response rate times the eligibility rate. The response rate for each index area is calculated from CE data over the most recent five year period, whereas the eligibility rate is the percent of addresses on the sampling frame with occupied housing units and is calculated using the most recent five years of data from the ACS, which also uses the Master Address File (MAF) as its frame. Since the response rates are different for the Interview and Diary Surveys, an optimization model is run for each survey. In Design 2010, the number of addresses is calculated annually using the most current response and eligibility rates.

A similar nonlinear optimization model was used in the previous design, with a subtle difference (King et al., 2008). The decision variable was the number of usable interviews, around 7,000, and not the number of addresses, which is 12,000. In the new design addresses are allocated instead of usable interviews. This change moves the nonresponse adjustment to an earlier step in the process. In the past a nonresponse adjustment was made to inflate the number of usable interviews up to the number of usable addresses that needed to be selected. Also, in the previous design, the sample size was determined once, and there were two linear constraints on the number of usable interviews: one for urban index areas and the second constraint for rural index areas.

Other updates to the sample design were considered through research projects conducted prior to the new sample design implementation. However, a decision was made to not include them because the results of the research did not provide enough evidence of improvement to the sample design. For example, one of the research projects suggested that cost savings could be obtained if the sample was clustered, where two, three or four neighbors would be in sample at the same time. However, it was concluded that even though there is some cost savings associated with clustering (Reyes-Morales et al., 2008) there would have to be an overall sample size increase to maintain the current variance on the key survey estimate due to the correlation between neighbors' expenditures which would require an increased budget (Ash et al., 2010).

3.2 New Sampling Frames and Sample Coordination

After determining the sample size for every PSU, the next step is selecting a sample of households in them and that requires sampling frames. The sampling frames for Design 2010 are new and are specially designed to meet the needs of the Demographic Household Surveys of the Census Bureau. The surveys have the same population of interest: the civilian non-institutionalized population of the United States and therefore able to share the same sampling frames and sampling systems. The new sampling frames are designed to meet the surveys requirement of sample coordination and allow a more frequent, survey-specific, sampling process.

In Design 2010 the Census Bureau has three sampling frames that are shared by all of its Demographic Household Surveys, including CE: the Unit, Group Quarters (GQ), and Coverage Improvement frames. All three frames are created from the Census Bureau's MAF, which is basically a list of all residential addresses identified in the 2010 census plus biannual updates from the U.S. Postal Service (Nguyen et al., 2011).

The Unit frame is the largest frame and it contains both existing housing units and new growth units. It has over 98% of the MAF's addresses. The GQ frame is also created from the MAF, but it is much smaller. It is a list of housing units that are owned or managed by organizations for residents who live in group arrangements such as college dormitories and retirement communities. The Coverage Improvement frame is also created from the MAF, but it is supplemented by additional housing units identified through address canvassing procedures. It contains housing units that are primarily in rural areas where there is a high concentration of non-city-style addresses[2], but CE decided not to use it.

In Design 2010, there is a major change in the updating method of the Unit Frame. Prior to Design 2010, the frame was a static list of addresses that was updated once per decade, but now it is a dynamic list of addresses that is updated twice per year with information from the Postal Service. That allows the sampling frequency to be increased from once per decade to once per year. The frame also allows for mid-year growth to be incorporated into the samples via an extension of the frame called a skeleton, a set of empty records, which are filled-in with new growth during the six month update of the frame. The skeleton is sampled during the regular annual sampling process using the same sampling rate as the Unit frame. The skeleton sample becomes active only when filled with new growth during the mid-year frame update.

Conversely, the GQ frame does not have a growth component and is updated every three years. If a new GQ is created after the frame creation, that GQ will not be included into the sample until the next frame creation process. If the size of a selected GQ changes, those changes are taken into account during the GQ sampling process.

In Design 2000, four frames represented the civilian non-institutional population: Unit, Group Quarters (GQ), Area, and Permit. Most addresses in the United States are covered by the Unit and GQ frame. The Unit frame is the largest frame and represents regular housing units. The GQ frame represents group living arrangements such as a college dormitory. The Permit and Area frames identified new addresses or new growth. The permit frame was a skeleton frame, a list of empty cells, which was filled in with new growth identified by building permit offices throughout the life of the design. The Area frame was used in locations with high concentrations of non-city-style addresses or no building permits were available and required a field listing procedure to capture new growth. These frames were created once, at the beginning of the design and the sample was selected for the next ten years.

The sample coordination between the Demographic Household Surveys was an easy implementation in Design 2000 because the sampling was done once. For Design 2010, the sample coordination is more complex because the sample selection is done annually.

---

[2] A non-city-style address is one whose format uses a rural route and box number, or a post office (PO) box, instead of a house number and a street name.

In order to achieve this sample coordination, the surveys enacted a set of common sampling rules and controls on the actual frames to facilitate this process. For example, the need to sample both births (new growth) and deaths (demolished units) during the sample selection process by all the surveys is a new sampling requirement. This is now necessary in the new design so that in the future, those units are sampled at the same rates as the existing units. Then once the sample is prepared for interviewing, their status is evaluated prior to being sent out for interview and at this time, the deaths are filtered out of the sample. Another frame issue is ensuring that once a survey selects a sample of housing units, the sample "resting period", or the 5-year time period needed between a households last scheduled interview and next possible selection for a new survey, is the same for all selected housing units and is independent of whether or not the household was actually sent out for interview. This ensures that the left over frame universe maintains its properties as an unbiased universe. For example, some surveys sample the frame at a higher rate and then implement a subsampling process in order to target specific populations. The sampling rules would force the initial sample to have the same resting period as the sample that was actually sent out for interview. To ensure that there is enough sample on the sampling frames for all surveys during the life of the sample design, all of the sampling fractions of the surveys are evaluated prior to each round of sample selection. During this evaluation, changes to every surveys sampling fraction could occur to reduce the amount of sample being selected from the frame, within a particular county. The limits are imposed on all surveys that are in the affected county, and these limits control the amount of sample that could be selected. Any adjustments are recorded and incorporated into the sample weights for each survey.

3.3 Within-PSU-Stratification

Even though the sampling frames are shared by all of the Demographic Household Surveys, each survey selects an efficient sample differently. The CE Survey orders the households on the sampling frame in such a way that when a systematic sample is selected, households from every economic stratum are well-represented in the survey. Households on the frame are sorted by variables whose values are known for every household on the frame and which are correlated with the surveys main variable of interest, the average annualized total expenditure per household on all items. Sorting the households this way has the effect of stratifying the frame and since the sorting procedure is done independently within each PSU, it is called "within-PSU-stratification."

CE draws its sample from two frames (Unit and GQ), but only the Unit frame uses a CE-specific variable to sort the households from poor-to-rich before drawing a sample of them. The GQ frame uses a generic variable common to all Census Bureau Demographic Household Surveys. In the Unit frame, the stratification variable (the sorting variable) is created from the number of occupants in each household, their housing tenure (owner/renter), and the market value of their home (for owners) or the rental value of their apartment or home (for renters). These variables are used because they are correlated with expenditures: households with more people tend to be wealthier than those with fewer people; homeowners tend to be wealthier than renters; and people living in high-price housing units tend to be wealthier than those living in low-price housing units.

The number of household occupants and their housing tenure come from the 2010 decennial census and are on the MAF, while monthly rental and property values come from the households surveyed by ACS and are on its 5-year data file. In Design 2010 the stratification variables are updated annually incorporating the most up-to-date ACS

estimates. Table 1 shows the Design 2010 within-PSU-stratification for geocoded addresses with complete tenure and vacancy information.

Table 1. Design 2010 Within-PSU Stratification Value Assignment

	Housing Value Quartile	*Number of Household Occupants*				
		1	2	0	3	4+
Estimated Monthly Rent for Renters(quartiles)	1	10	11	12	13	14
	2	25	24	23	22	21
	3	30	31	32	33	34
	4	45	44	43	42	41
Estimated Market Value of Home for Homeowners (quartiles)	1	50	51	52	53	54
	2	65	64	63	62	61
	3	70	71	72	73	74
	4	85	84	83	82	81

The monthly rental and property values are aggregated into four quartiles, which are defined separately by county using data collected by ACS. The Census Bureau partitions every county into a large number of "blocks," and then CE staff aggregates those blocks into a small number of contiguous geographic "domains" having 50-100 renters who were in the ACS survey. Their median rental value is then computed using their ACS data and the median value is assigned to every household in the domain that reported being a renter in the 2010 census. The process generated a few dozen geographic domains per county, each of which had its own median rental value, and then quartiles were formed by stratifying the domains into four groups. Then the process was repeated for homeowners.

In Table 1, all of the renters are at one end of the stratification and all of the owners are at the other end of the stratification. The renters and owners are subdivided into quartiles because monthly rental and property values vary by geographic area and quartiles provide a more equal distribution of the addresses than raw dollar amounts. Vacant housing units are put in the middle column because although they were vacant at the time of the decennial census, when CE's field representatives visit them they could be in any of the four non-zero categories. The serpentine sorting order guarantees a good mixture of expenditure levels in the sample. This makes sample selection efficient for the CE surveys and minimizes the variance in the second-stage.

The within-PSU-stratification variable for the Design 2000 Unit frame was similar to the Design 2010 stratification variable described above, but their data came from different sources. In Design 2000, the number of occupants and tenure came from the 2000 decennial census short form, while the rental and property value came from its long form. In Design 2010, the number of occupants and their tenure still came from the decennial census, but since the long form was discontinued the rental and property value was taken from ACS (Steinberg et al., 2009). Also, in Design 2000, vacant units (0 occupants) were placed in the leftmost column instead of the middle column because 0 normally comes before 1, 2, 3, and 4; and the rows alternated between renters and owners, placing poor renters next to poor homeowners to keep poor people together. Similarly, rich renters were placed next to rich homeowners to keep rich people together. However, research showed that renters tend to be uniformly poorer than homeowners (the richest renters are

poorer than the poorest homeowners), which led to a decision to completely separate the renters from the owners in Design 2010 (Lineback et al., 2009).

The within-PSU-stratification variable used in the GQ frame is pre-defined and not unique for each survey. It uses a geographic and block level sort on "percent of college housing." The college housing population is very different than the rest of the GQ population (Jonas et al., 2012), so using it as the within-PSU-stratification variable produces a more representative systematic sample of GQ housing. For Design 2010, the GQ frame is re-created every three years and at that time any newly discovered GQs will be included in the next round of GQ sampling (Nguyen et al., 2011). By contrast, in the previous design, the GQ sample was selected for the entire decade at the beginning of the sample design.

3.4 Selecting a Systematic Sample of Households

The Interview and Diary households are selected jointly, in one sample selection process for each frame. The GQ frame sampling selects three years of sample in one round of sampling, and the unit frame sampling selects enough sample for one year. The sample sizes for the combined selection are created by first taking the larger sample size generated by the optimization program described in Section 3.2. The larger sample size for the PSU from either the Diary Survey or Interview Survey is doubled to ensure that enough sample is selected for both surveys. The selection is planned such that alternating sample units are used in the Interview Survey or Diary Survey, and to achieve the survey specific sample sizes, a sample reduction process is planned to randomly remove housing units from the survey which required the smaller sample.

Each county has its own sample selection process. Once the list of housing units within a county are sorted using the within-PSU-stratification, the first housing unit is randomly selected using a dependent random number generator. The dependent random number generator is used in the sample selection process to ensure that the randomness introduced by the number generator does not affect the overall desired sample size. Then the remaining housing units are selected by taking every k^{th} housing unit on the ordered list. The number k is the sampling interval for the county and it is computed independently for each PSU by dividing the total number of housing units from the MAF by the desired sample size.

The effects of the sample coordination of the CE sample with the other household surveys could also affect the sample selection process if a particular county that CE selects sample from is flagged as being "crowded." The term "crowded" identifies a county in which the combined survey sampling rate, across all surveys, for that particular county was identified to be too much for the county to handle. Once a county is flagged, the sampling rates allowed for that county are capped for all surveys to ensure that there are enough housing units for all the surveys to sample from. These adjustments to the sampling rates are rare but would affect the overall sample sizes at the PSU level for all the coordinated sample surveys.

4. Sample Administration and Maintenance

The last part of the second-stage sample design is the planning that occurs after sample selection. Each survey has its own method of planning how each sampled housing unit will enter the interview process and how the new design will be introduced into the current interview cycle.

Sample coding is the process of assigning each housing unit in the sample to either the Diary or the Interview Survey. The housing units are labeled with sample codes to identify their assigned survey. The sample coding process also assigns the housing units to: (1) a time frame for interviewing; (2) half-samples, which are used in variance estimation; and (3) sample reduction codes.

When assigning the sample codes, it is necessary to order the selected units, called hits, by original sort order. The goal of the code assignments is for each separate sample code and sample code combination to be a subsample of the overall systematic random sample. Furthermore, the subsample must be a systematic random sample with hits that are equi-distant to each other (Ash 2011).

Sample designations are sample codes that identify whether a housing unit is assigned to the Diary Survey or the Interview Survey. Sample designations also indicate if a housing unit is a production unit or a reserve unit. Reserve units are supplemental housing units that are set aside for special research projects. All other housing units are called production units and are a part of the main sample. The four sample designations are Interview Production (Q), Interview Reserve (X), Diary Production (D), and Diary Reserve (E). In the previous design, both the production and reserve sample designations were in the same hit string. However, in Design 2010, the Diary reserve sample is included with the Interview production sample and the Interview reserve sample is included with the Diary production sample. This structure ensures the reserve sample for either the Diary or Interview Survey will not be geographically close to its production sample, if it is used (hits are geographically close due to the sort order). A number is appended to the sample designation to indicate the year in which the sample was selected.

For the Diary Survey, the Diary Placement Day is the earliest day of the year when the diary is to be placed. This is determined by uniformly assigning the sample codes quarter, week, and day. For the Interview Survey, the interview dates are determined from two sample codes called panel and rotation. The rotation sample code is the quarter of the year when the sample designation is introduced. The panel represents the month of the quarter when the sample units are interviewed. There are several other sample codes of lesser importance such as reduction groups and half-samples. Reduction groups are numbers between 1 and 101 assigned to every household in the sample that are used to reduce the sample. To reduce the sample by 1%, a reduction code is randomly selected, and units with that reduction code are excluded from the sample. The half-sample code is a special sample code that splits the sample into equally sized groups and is used in estimating the variance.

The sample codes are systematically assigned after sorting the housing units in a specific order. For example, in assigning the sample designations, the file is first sorted by the original hit order, and then housing units are sequentially assigned to the following samples: Diary Production, Interview Reserve, Interview Production, and Diary Reserve. The other sample codes are assigned similarly but with different sort orders. The sort order is important to prevent correlations from being generated between some of the coded variables. For example, the housing units are sorted to avoid assigning all the odd numbered half samples to the same quarter, or the even numbered half samples to the same panel group.

The Design 2010 Diary Survey sample was introduced in January 2015, and the Interview Survey sample was gradually phased-in over the eleven-month period of February through December 2015.

5. Other Changes and Summary

In Design 2010, there were improvements to the frames and timing of the sample selection process. Although not specific to CE, these changes are briefly discussed.

The MAF was updated with Global Positioning System (GPS) coordinates that were collected during Decennial 2010 address canvassing operations. Most of the addresses on the MAF have GPS coordinates (94 percent) and these GPS coordinates will be passed to field representatives as an additional method to use when trying to locate their case assignments (Winstead et al., 2011).

The coordination of the Demographic Household Surveys into one on-going sample selection system has the added bonus of being able to in-activate and re-activate sample units after a pre-determined resting period that is specific to each survey. This allows a sample unit that has already been selected for interview, to have a pre-defined resting period which prevents the unit from being selected again within that time period (Nguyen et al., 2011).

As a summary, Table 2 provides a quick reference to highlight some of the changes between the old and new design discussed in the previous sections.

Table 2. Design 2000 vs Design 2010 for the CE Surveys

Sample Design Element	**Design 2000 Details**	**Design 2010 Details**
PSU Selection Frequency	Every 10 years	Every 10 years
PSU Name		
1st letter	A, X, Y, Z	S, N, R
2nd letter	Census Region	Census Region
3rd letter	3rd and 4th digits are	Census Division
4th letter	Stratum Indicators	Stratum Indicator
First Stage PSUs	75 non rural PSUs 16 rural PSUs	75 non rural PSUs 16 rural PSUs
Second Stage Frames	4 Frames: Unit, Area, Permit, GQ	2 Frames: Unit, GQ
Second Stage Stratification Clusters	41 Strata	47 Strata
New Growth	Area, Permit: *ongoing*	Unit: *every 6 months* GQ: *every 3 years*
Frame Creation and Second Stage Sampling Frequency	Every 10 years	Unit Frame: Yearly GQ Frame: Every 3 Years

6. Future Research

CE plans a major revision to both the Interview and Diary Surveys in Design 2020. The proposed design includes two waves of data collection twelve months apart. The two surveys will combine and the same household will participate in both waves. Each wave is composed of two visits with a household member serving as a respondent. The first visit is an in-person interview in which the field representative collects easily recalled expenditures from the previous three months. The field representative will ask the respondent to collect records for expenditures such as utilities for the three month period prior to the second interview. Also, on the first visit, the field representative will train all eligible household members on using the electronic diary, which individual expenditures will be entered for the next week. During the second interview, which occurs one week after the first visit, the diaries will be reviewed for missed expenditures and then large expenditures from requested records at the first interview will be recorded. Twelve months later, the process will be repeated with the same interview structure. Hopefully, the new design change will increase response rates by reducing respondent burden. The new design will alleviate the repetitive collection of some expenditure like mortgage payments which do not change from month to month. One of the downsides of the new design is that four continuous quarters of data from the same household will not be available for research projects. The new design will have minimal impact on the sample selection procedures discussed in this paper.

7. Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Census Bureau and the U.S. Bureau of Labor Statistics.

8. References

Ash, S., Rawlings, A., Swanson, D., Reyes-Morales, S., and Chopova, B. (2010). U.S. Census Bureau Memorandum "Consumer Expenditure Survey 2010 Redesign Research: Cluster Sampling Research for the Consumer Expenditures Surveys" From Killion, R., to Ryan, J., dated March 30, 2010.

Ash, S. (2011). U.S. Census Bureau Memorandum "High-Level Requirements for the Sample Code Assignments for the Consumer Expenditures Diary and Quarterly Interview Surveys and the National Crime Victimization Survey, version 1.0", dated November 28, 2011.

Ernst, L. (1986). Maximizing the overlap between surveys when information is incomplete. European Journal of Operational Research, Vol. 27, 192 - 200.

Ernst, L., Johnson, W., and Larson, W. (2007). Reduction of the Sample Areas in the Consumer Price Index and Consumer Expenditure Survey Designs. Proceedings of the American Statistical Association, Section on Survey Methods Research, 2918-2924

Johnson, W., Paben, S., and Schilp, J. (2012). Use of the Sample Overlap Methods in the Consumer Price Index Redesign. 2012 Proceedings of the International Conference on Establishment Surveys.

Jonas, K., and Kuwik, C. (2012). U.S. Census Bureau Memorandum "2010 Sample Redesign Requirements for GQ Within-PSU Sampling (Doc. #2010-3.4-Q-3, version 1.01)" from Farber, J. to Williams, T., dated October 4, 2012.

King, S. and Johnson-Herring, S. (2008). Selecting a Sample of Households for the Consumer Expenditure Survey. Consumer Expenditure Survey Anthology, 14-19.

King, S., Schilp, J., and Bergmann, E. (2011). Assigning PSUs to a Stratification PSU. Proceedings of the American Statistical Association, Section on Survey Methods Research, 2235-2246.

King, S. (2012). A "First-Cut" at Forming Rural PSUs. Proceedings of the American Statistical Association, Section on Survey Methods Research, 4055 -4066.

Lineback, J., Rawlings, A. and Steinberg, B. (2009). U.S. Census Bureau Memorandum "Consumer Expenditure Survey 2010 Redesign Research: Imputation of property value and rent for within-Primary Sampling Unit stratification" January 9, 2009, from Lineback, J., Rawlings, A. and Steinberg, B. to The Record, dated February 10, 2009.

Murphy, P., (2008). An Overview of Primary Sampling Units (PSUs) in Multi-Stage Samples for Demographic Surveys. Proceedings of the American Statistical Association, Section on Government Statistics, 2856-2863.

Nguyen, T., and Gerstein, A., (2011). Sample Design Research in the 2010 Sample Redesign. Proceedings of the American Statistical Association, Section on Survey Methods Research, 4359-4372.

Perkins, W. (1970). 1970 CPS Redesign: Proposed Method for Deriving Sample PSU Selection Probabilities Within 1970 NSR Strata. Memo to Waksberg, J., U.S. Bureau of the Census.

Reyes-Morales, S., Davis, J., Chopova, B., Ash, S., Lineback, F., and Rawlings, A. (2008). U.S Census Bureau Memorandum "Cluster Sampling Research Team Final Report" From Cluster Sampling Research team to Killion, R.A., Ryan, J., Henderson, S., dated June 27, 2008.

Ryan, J., (2013). Bureau of Labor Statistics Memorandum "Bounding Interview Elimination" From Ryan, J. to Pickering, C., dated April 2, 2013.

Swanson, D., Reyes-Morales, S., Ash, S., and Croos, J. (2012). U.S. Census Bureau Memorandum "Consumer Expenditure Survey 2010 Redesign Research: Sample Allocation Research for the Consumer Expenditure Interview Survey" From Killion, R., to Ryan, J., dated March 20, 2012.

Swanson, D., Reyes-Morales, S., Steinberg, B., Ash, S., and Dumbacher, B. (2011). U.S. Census Bureau Memorandum "Consumer Expenditure Surveys 2010 Redesign Research: Optimizing the Sample Allocation for the Consumer Expenditure Surveys" From Killion, R., to Ryan, J., dated August 1, 2011.

Steinberg, B., Ash, S., Flores, L., Rawlings, A., and Dompreh, I. (2009). U.S. Census Bureau Memorandum "Consumer Expenditure Survey 2010 Redesign Research: Determining the Within Primary Sampling Unit Stratification" From Killion, R., to Ryan, J., dated December 8, 2009.

Winstead, A., Byrne, R., Kuwik, C., and Harclerode, S. (2011). How the U.S. Census Bureau's Redesign of Household Survey Sampling Impacts Field and Clerical Operations. Proceedings of the American Statistical Association, Section on Survey Research Methods, 3880-3889.