

Consumer Expenditures and Income: *Overview*

The **Consumer Expenditure Surveys (CE)** are nationwide household surveys conducted by the U.S. Bureau of Labor Statistics (BLS) to study how U.S. consumers spend their money. The surveys are the only federal government data collection effort to obtain information on the complete range of consumers' expenditures, income, and demographic characteristics, in the same survey, directly from consumers. BLS publishes 12-month estimates of consumer expenditures annually, with the estimates summarized by various income levels and demographic characteristics. BLS also produces annual public use microdata files and an online database to help researchers analyze the data in more detail.

The CE consists of two separate surveys, the Interview Survey and the Diary Survey. BLS designs the Interview Survey to collect data on large and/or recurring expenditures that respondents can be expected to recall for a period of 3 months or longer. In general, expenditures reported in the Interview Survey are either relatively large, such as those for property, automobiles, and major appliances, or recur on a regular basis, such as for rent, utilities, or insurance. BLS designed the Diary Survey to collect data on frequently purchased items, which can be difficult to recall even a few weeks later. These items include food and beverage expenditures at home and in eating places; housekeeping supplies and services; nonprescription drugs; and most personal care products and services. Together, the data from the two surveys cover the complete range of consumer unit expenditures. The U.S. Census Bureau collects CE data for BLS.

CE data are unique in enabling data users to examine the association of expenditures and income of consumers to consumer unit characteristics. CE data are of value to government and private agencies interested in studying particular segments of the population, such as the elderly, low-income families, urban families, and those receiving Supplemental Nutrition Assistance Program food benefits. Economic policymakers use CE data to analyze the effects of policy changes on living standards across diverse socioeconomic groups, and econometricians find the data useful in constructing models to estimate economic outcomes. Market researchers find consumer expenditure data valuable in analyzing the demand for different goods and services. BLS uses the survey data to produce the Supplemental Poverty Measure (SPM) thresholds, which in turn are used by the Census Bureau to produce SPM statistics, and by the Bureau of Economic Analysis in revising its benchmark estimates of selected items in the expenditure and income components of the national accounts. The U.S. Department of Agriculture uses CE information to estimate the cost of raising a child from birth to age 18, and the Internal Revenue Service uses expenditure data from CE to calculate alternate sales tax standard deductions. In addition, the Department of Defense uses the data in determining cost-of-living allowances for military personnel living off military bases.

Another primary reason BLS undertakes the CE is to provide weights for the Consumer Price Index (CPI). That is, results of the CE are used to estimate upper-level spending weights for the Chained CPI for All Urban Consumers (C-CPI-U, monthly) and the CPI-U and CPI-W (annually) indexes as well as to the sampling weights for lower-level index calculations; to calculate item sampling probabilities (annual); and to derive outlet sample frame and selection probabilities for the CPI Commodities and Services Survey (semi-annual). In general, the CE provides expenditures at the lowest index level and then they are adjusted and aggregated to calculate data for CPI uses.

Quick Facts: Consumer Expenditures and Income	
Subject areas	Income, Consumer spending , Pay
Key measures	<ul style="list-style-type: none">• Aggregate expenditures• Average expenditures• Expenditure shares

Quick Facts: Consumer Expenditures and Income	
How the data are obtained	Survey of households
Classification	Demographic
Periodicity of data availability	Semi-annual, Annual
Geographic detail	Metro area, Census region, National
Scope	Civilian noninstitutional population
Sample sizes	20,000 independent interview surveys and 11,000 independent diary surveys completed annually
Reference Period	Calendar year
Revision Information	Data are not routinely revised.
Key products	<ul style="list-style-type: none"> • Data tables • Public use microdata • Publications • LABSTAT database • Research products
Program webpage	www.bls.gov/cex

Concepts

The Consumer Expenditure Surveys (CE) gather data on household expenditures, income, changes in assets and liabilities, and demographic characteristics of consumers in the United States. A consumer unit (CU) is the measurement unit collected for the eligible individuals represented in the expenditure reports.

The CU is defined as

- all members of a particular housing unit who are related by blood, marriage, adoption, or some other legal arrangement, such as foster children;
- a person living alone or sharing a household with others, or living as a roomer in a private home, lodging house, or in permanent living quarters in a hotel or motel, but who makes independent financial decisions;¹ or
- two or more unrelated persons living together who pool their income to make joint expenditure decisions.

Students living in university-sponsored housing are also included in the sample as separate CUs. Information on members living in the CU is identified by their relationship to the reference person, who is

defined as the first member mentioned by the respondent when asked to "start with the name of the person or one of the persons who owns or rents the home." Although sometimes used interchangeably, households and CUs are not the same; some households contain more than one CU.² In publications, and with CE respondents, "household" is occasionally used for simplicity, but nevertheless refers to the CU. Survey participants report dollar amounts for goods and services purchased by any member of the CU during the reporting period, regardless of whether payment was made at the time of purchase. The Bureau of Labor Statistics (BLS) defines expenditures as the actual financial obligation incurred for goods or services acquired by the CU from a source outside the CU at the time of acquisition. Expenditure amounts for items purchased by the CU include all applicable sales and excise taxes. Excluded from expenditure total amounts are any business-related expenditures (e.g., travel, lodging, etc.) and expenditures for which the CU is reimbursed (e.g., reimbursed medical expenses).

The CE is not a consumption survey. Consumption refers to everything a CU consumes, while consumer expenditures are limited to everything a CU purchases. For example, an item purchased by someone outside the CU, and given to the CU to consume, would be considered consumption and not an expenditure. Similarly, an item made by a CU member, but not directly purchased, would also be considered consumption and not an expenditure. Other examples of nonexpenditure consumption include in-kind benefits, service flows from durable goods, home production, and barter. These examples demonstrate that the CU does not have to make any explicit expenditures for consumption to occur, which implies there are not always records of what a CU consumes. As a result, consumption at a household or individual level is inherently difficult to measure.

NOTES

¹ Financial independence is determined by spending behavior regarding the three major expense categories: housing, food, and other living expenses. To be considered financially independent, the respondent must provide at least two of the three major expenditure categories, either entirely or in part.

² For more information of the differences between households and CUs, see <https://www.bls.gov/opub/mlr/2021/article/consumer-expenditure-survey-methods-symposium-and-microdata-users-workshop-2020.htm>.

Collections & Data Sources

The Bureau of Labor Statistics (BLS) obtains Consumer Expenditure Surveys (CE) Interview Survey and Diary Survey data by interviewing respondents about their expenditures, income, and characteristics. The U.S. Census Bureau selects the samples of household addresses and collects the data under contract with BLS.

Survey notification and collection method

A selected sample household address is mailed an advance letter from the Census Bureau informing the residents about the purpose of the survey and the upcoming visit by the interviewer. The Census Bureau conducts both the Interview Survey and the Diary Survey primarily by personal visit with some telephone interviewing. The interviewer uses a structured questionnaire to collect the demographic and income data in the Interview and Diary Surveys. The structured questionnaire is also used to collect expenditure data in the Interview Survey. Expenditure data in the Diary Survey are entered by the respondent into a paper or online diary form. Any eligible household member who is at least 16 years old can serve as the respondent in either survey.

Interview Survey details

The Interview Survey collects detailed data on types of expenditures that are usually fairly easy to recall for periods of 3 months or longer. On average, it takes 67 minutes to complete the interview.

Each consumer unit (CU) at a sampled Interview Survey address remains in the sample for four quarters and is interviewed every 3 months. The sample of addresses for each quarter is divided evenly across three monthly panels and each address remains in the same monthly panel each quarter. For example, if an address is first included in the sample in the second month of a calendar quarter (e.g., February), then it will be in the sample in the second month of the following three quarters (e.g., May, August, November). Because the sample is based on address, the CU that is interviewed at the address may be interviewed up to four times. However, if the CU moves from the sampled address, that CU will no longer be interviewed. Any new CU moving into the address will be interviewed instead for whatever time remains for that address in sample.

After the fourth interview, the address is dropped from the survey and replaced by a new address. For the survey as a whole, 25 percent of the sample in each quarter consists of new addresses introduced into the sample to replace the addresses that have been in the sample for four quarters. Data collected in each quarter are treated independently, so that published 12-month estimates are not dependent upon a particular family participating in the survey for a full four quarters.

Exhibit 1 shows how Interview Survey addresses rotate in and out of the sample. In this example, the first interviews start in April, May, and June 2021. Three months later, the second interviews begin. A CU at a sample address first interviewed in April 2021 is re-interviewed in July 2021, October 2021, and January 2021. And while the second set of interviews begins in July 2021 for the units first interviewed in April, a new set of addresses is starting their set of four interviews.

Exhibit 1. Quarterly Interview Survey rotation

Interview year and month		Interview set			
		1	2	3	4
2021	APR	a			
	MAY	b			
	JUN	c			
	JUL	d	a		
	AUG	e	b		
	SEPT	f	c		
	OCT		d	a	
	NOV		e	b	
	DEC		f	c	
2022	JAN			d	a
	FEB			e	b
	MAR			f	c

Exhibit 1. Quarterly Interview Survey rotation

Interview year and month		Interview set			
		1	2	3	4
	APR				d
	MAY				e
	JUN				f
	JUL				
	AUG				
	SEPT				

Note: The column headings are the interview numbers. Each letter designates a panel or group of household addresses. Each panel consists of interviews conducted every three months for four quarters.
Source: Bureau of Labor Statistics.

During the first interview (interview set 1 in exhibit 1) information is collected on demographic and family characteristics and on the inventory of major durable goods for each CU. This socioeconomic information is used by BLS to classify the CU for the publication of statistical tables and economic analyses. Also collected in the first and fourth interviews are wage, salary, employment, and income (including unemployment compensation; income from royalties, dividends, and estates; and alimony and child support) information of each member of a CU who is age 14 years or older.

Expenditure data are collected in each interview via multiple question patterns depending on the types of expenditures collected. One question pattern asks the respondent for the month of purchase of each reported expenditure. A second question pattern asks for quarterly amounts of expenditures. A third question pattern asks for the payment frequency and the amount based on said frequency. In the fourth interview, an annual supplement collects information on financial assets and liabilities (debts).

Diary Survey details

The Diary Survey is a separate survey designed to collect smaller, more frequently purchased items. Two separate instruments are used by the Census Bureau to collect Diary Survey data: a household characteristics questionnaire and a record of daily expenses (also known as the "diary"). In the household characteristics questionnaire, the interviewer records information pertaining to age, sex, race, marital status, and family composition, as well as information on the work experience and earnings of each member of the CU. Like the Interview Survey, BLS uses this socioeconomic information to classify the CU for publication of statistical tables, as well as for economic analyses.

The record of daily expenses is designed as a self-reporting diary, in which respondents record a detailed description of all expenses made by all CU members for two consecutive 1-week periods. Diary keeping begins on the day following the completion of the household characteristics questionnaire. Data collected each week are treated as statistically independent for publications—each week's diary is separately weighted to be representative of the sample. The diary is divided by day of purchase and by four

classifications of goods and services—meals, snacks, and drinks away from home; food and drinks for home consumption; clothing, shoes, jewelry, and accessories; and all other products, services, and expenses. The items reported are subsequently coded by the Census Bureau so that BLS can aggregate individual purchases for representation in the CPI and for presentation in statistical tables.

The Diary Survey asks for almost all expenses that the CU incurs during the survey week. Expenses incurred by CU members while away from home overnight or longer are not in scope for the diary and are excluded. It takes approximately 20 minutes over two visits for the interviewer to collect the demographic data and to instruct the respondent on how to keep the diary. It is estimated that it takes the respondent 10 minutes each day to enter expenses into the diary.

Census quality control and confidentiality

Data collection quality control and data integrity are provided by a re-interview program, which evaluates the performance of the individual interviewer, to determine how well the procedures are being carried out in the field. The re-interview is conducted by a Census Bureau supervisor or an interviewer at a national processing center data contact center (a centralized telephone call center for conducting interviews). Subsamples of approximately 9 percent of households in both the Interview and Diary Surveys are re-interviewed on an ongoing basis.

All data collected in both surveys are subject to Census Bureau and BLS confidentiality requirements that prevent the disclosure of the respondents' identities. The information that respondents provide is used solely for statistical purposes. All Census Bureau and BLS employees who work with the CE data take an oath of confidentiality and are subject to fines and imprisonment for improperly disclosing information provided by respondents. Confidentiality certification training is required annually.

Names and addresses are removed from all forms and datasets prior to transmission from the Census Bureau to BLS and are not included in any statistical releases. At BLS, the data are processed and stored on secure servers, with access limited to employees having the appropriate security clearances. As a further precaution, BLS applies certain restrictions to the microdata available on the public-use files. These include geographical and value restrictions that prevent the identification of respondents.

Data collection and Census Bureau processing

The Census Bureau, under contract with BLS, carries out data collection for both the Interview Survey and the Diary Survey. In addition to its collection duties, the Census Bureau does field editing and coding, checks consistency, ensures quality control, and securely transmits the data to BLS. In preparing the data for analysis and publication, BLS performs additional review and editing procedures.

Interview Survey

The first step of quality control in the Interview Survey is in the computer assisted personal interview (CAPI) instrument that Census Bureau interviewers have used to collect interview survey data since April 2003. The CAPI instrument enforces question skip patterns and prompts the interviewer to confirm unusually low or high expenditure values. For some expenditures, the CAPI instrument also provides a list of expenses reported in previous interviews to prevent duplication of reports.

At the completion of the interview, data are electronically transferred from an interviewer's laptop to the Census Bureau's master control system. The Census Bureau conducts post-processing by reformatting the data into datasets based on the required BLS output structure and performing special processing, including converting missing values to special characters and removing any personally identifiable information. Some "inventory" data that are not expected to change, such as vehicle and mortgage information, are copied into an input file that is loaded onto the laptops for subsequent interviews during the next quarter. This way, a few questions related to those expenditures are updated each quarter, rather than an entire data record. Names and addresses of respondents are not transmitted to BLS.

Diary Survey

At the beginning of the 2-week collection period, the Census Bureau interviewer, using the household characteristics questionnaire (a CAPI instrument), records demographic information on members of each sampled CU. In some cases, interviewers will also ask respondents at this time about work experience and income. Upon completion of the household characteristics questionnaire, the interviewer provides the respondent with two 1-week diaries and instructs them how to fill it out. During the diary-keeping period, the interviewer periodically checks in with the CU to answer any questions. At the end of the diary keeping period, the interviewer collects the diary, reviews the entries, and works with the CU to add any expenses that may have been missed. During this time, the interviewer will ask questions about work experience and income, if they have not already done so.

Data from the household characteristics questionnaire are reviewed for completeness and consistency and then are scanned and transcribed into a database. Expenditure descriptions written by the respondent are assigned to different codes by staff with the assistance of a computer program developed to assign codes for commonly used descriptions. The final data are transmitted to Census Bureau's headquarters, along with any scanned image files of the diaries. Census Bureau staff combines expenditure data from diaries with data collected in the household characteristics questionnaire, removes personal identifiable information, and transmits the merged files to BLS on a monthly basis.

Questionnaire revisions

BLS periodically provides new requirements to the Census Bureau for updating the Interview Survey questionnaire and the Diary Survey form to incorporate new products and services, clarify instructions, improve instrument navigation, incorporate changes requested by stakeholders, and streamline the interview by deleting outdated items. Although major changes to the questionnaire are made every other year, BLS staff who work on the CE continuously monitor the emergence of new goods and services available in the marketplace, as well as changes in the relative importance of existing items in consumers' budgets. Updated information on how to report new goods and services is provided to the interviewers on a regular basis.

Sample Design

The Consumer Expenditure Surveys (CE) is a nationwide household survey representing the entire U.S. civilian noninstitutional population. It includes people living in houses, condominiums, apartments, and group quarters such as college dormitories. It excludes military personnel living overseas or on base, nursing home residents, and people in prisons. The civilian noninstitutional population represents more than 98 percent of the total U.S. population.

Selection of households

On behalf of the Bureau of Labor Statistics (BLS), the U.S. Census Bureau collects CE data from a representative sample of households across the United States. The Census Bureau collects data for the Interview Survey and the Diary Survey. BLS first draws the primary sampling units (PSUs) for the Interview Survey and the Diary Surveys. Subsequently, the Census Bureau draws a random sample of households inside those geographic areas.

The selection of households for the survey begins with the definition and selection of PSUs. BLS designs PSUs to be small clusters of counties grouped together into geographic entities called "core-based statistical areas" (CBSAs), which are defined by the Office of Management and Budget (OMB) for use by federal statistical agencies in collecting, tabulating, and publishing federal statistics. BLS uses OMB definitions from 2012 for the CE. There are two types of CBSAs: metropolitan and micropolitan. Metropolitan CBSAs are areas that have an urban "core" of 50,000 or more people, plus the adjacent counties that have a high degree of social and economic integration with the core as measured by commuting ties. Micropolitan CBSAs are similar to metropolitan CBSAs but they have an urban core of

10,000 to 50,000 people. Areas without an urban core or whose urban core is under 10,000 people are called non-CBSA areas. These non-CBSA areas are considered rural PSUs and are defined by BLS rather than by OMB. See a [complete list of CBSAs](#) in the United States and a detailed description of how they are defined.

Starting in 2015, the geographic sample used in the survey consists of 91 PSUs that are classified into three categories based on the three PSU types described above and their populations in the 2010 decennial census:

- Self-representing PSUs (known as "S" PSUs) are those selected with certainty. There are 23 "S" PSUs, which are metropolitan CBSAs with a population over 2.5 million people.
- Non-self-representing PSUs (known as "N" PSUs) are those selected randomly. There are 52 "N" PSUs, which are metropolitan and micropolitan CBSAs with a population under 2.5 million people.
- Rural PSUs (known as "R" PSUs). There are 16 "R" PSUs, which are non-CBSA areas.

The 23 "S" PSUs are the largest CBSAs in the country, and they were selected with certainty for the CE sample. The 52 "N" and 16 "R" PSUs are smaller CBSAs and non-CBSA areas that were randomly selected from the rest of the country, with their probabilities of selection being proportional to their populations. BLS uses all 91 of these PSUs for the CE sample. The Consumer Price Index (CPI) also uses these PSUs in its sample, except it uses only the 23 "S" and 52 "N" PSUs in its sample. The CPI does not use the 16 "R" PSUs because it measures inflation only in urban areas of the country.

Within these 91 PSUs, the list of addresses from which the sample is drawn comes from two sources called "sampling frames." The primary sampling frame for both the Diary Survey and the Interview Survey is the Census Bureau's master address file (MAF). That file has all residential addresses identified in the 2010 census and is updated twice per year with the U.S. Postal Service's delivery sequence file. Over 99 percent of the addresses used in the survey come from the MAF. It is supplemented by a small group quarters frame, which 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. Less than 1 percent of the addresses used in the CE come from the group quarters frame.

The Interview Survey is a rotating panel survey in which approximately 13,000 addresses are contacted each calendar quarter of the year for the survey. One-fourth of the addresses that are contacted each quarter are new to the survey. Usable interviews are obtained from approximately 5,000 households at those addresses each quarter of the year. After a housing unit has been in the sample for four consecutive quarters, it is dropped from the survey, and a new address is selected to replace it.

The Census Bureau selects a sample of approximately 18,000 addresses per year from the MAF and group quarters frames to participate in the Diary Survey. Usable diaries (two 1-week diaries per household) are obtained from approximately 6,700 households at those addresses. Diaries are not obtained from the other addresses because of refusals, vacancies, ineligibility, or the nonexistence of a housing unit at the selected address. The placement of diaries is spread equally over all 52 weeks of the year. BLS increased the sample sizes in 2020 to obtain sufficient outlet information for the CPI sample design, which previously used a different method of obtaining the information. Prior to 2020, the CE's sample sizes were 12,000 addresses per calendar quarter in the Interview Survey, and 12,000 addresses per year in the Diary Survey.

Cooperation levels

Response data for the 2019 CE are shown in table 1. For the Interview Survey, each unique CU provides up to four usable interviews per year. For the Diary Survey, each unique CU provides up to two usable diaries (weeks 1 and 2). Most Diary Survey respondents participate in both weeks.

There are three general categories of nonresponse:

- *Type A nonresponses* are refusals, temporary absences, and noncontacts.
- *Type B nonresponses* are vacant housing units, housing units with temporary residents, and housing units under construction.
- *Type C nonresponses* are nonresidential addresses, such as destroyed or abandoned housing units, and housing units converted to nonresidential use.

Response rates are defined as the percentage of eligible housing units (that is, the designated sample less type B and type C nonresponses) from which usable interviews are collected by the Census Bureau. In the 2019 Interview Survey, there were 40,389 eligible housing units from which 21,701 usable interviews were collected, resulting in a response rate of 53.7 percent. In the 2019 Diary Survey, there were 20,244 eligible housing units from which 10,682 usable interviews were collected, resulting in a response rate of 52.8 percent. (See table 1.)

Table 1. Analysis of response in the Consumer Expenditure Surveys, 2019

Sample unit	Interview Survey	Diary Survey
Independent addresses designated for survey	47,799	24,566 ¹
Less: type B and type C nonresponses	7,410	4,322
Equals: eligible units	40,389	20,244
Less: type A nonresponses	18,688	9,562
Equals: interviewed units	21,701	10,682
Percentage of eligible units interviewed	53.7	52.8

¹ The number of Diary Survey addresses (12,283) multiplied by two weekly diaries.

Note: Type A nonresponses are refusals, temporary absences, and noncontacts. Type B nonresponses are vacant housing units, housing units with temporary residents, and housing units under construction. Type C nonresponses are nonresidential addresses, such as destroyed or abandoned housing units, and housing units converted to nonresidential use.

Source: U.S. Bureau of Labor Statistics.

Calculation

The Bureau of Labor Statistics (BLS) processes and prepares Consumer Expenditures Survey (CE) microdata for analysis and publication. At a high level, this processing includes ensuring consistency among reported values, identifying and correcting errors (e.g., misclassified expenditures) in the data, imputing missing values (see below), and classifying (or “mapping”) expenditures to BLS spending categories. In addition, the primary statistic calculated is the average annual expenditure per consumer

unit (CU). It is a weighted average whose calculation follows well-established statistical principles. BLS computes weights to allow the sample data results to reflect the population, measured in CUs. In addition, BLS adjusts data by adding sales tax, netting out reimbursements, and excluding business-related expenses.

Data adjustment by survey

BLS adjusts data to two surveys: Interview Survey and Diary Survey.

Interview Survey

BLS completes three major types of CE data adjustment routines: imputation; allocation; and time adjustment. *Imputation* routines are used for income tax estimation, and to “fill in” or correct missing or invalid entries. Imputation addresses all types of the data (demographics, income, and expenditures) except assets. *Allocation* routines are used for respondents who provide insufficient detail to meet tabulation requirements. For example, combined expenditures for the fuels and utilities group are allocated among the components of that group, such as natural gas and electricity. *Time adjustment* routines are used to classify expenditures reported quarterly by month of occurrence, prior to aggregation of the data to calendar-year expenditures.

Diary Survey

Two types of data adjustment routines, imputation and allocation, improve the Diary Survey estimates. BLS imputes missing attributes, such as age, sex, or expenditure amount. Allocation routines transform reports of nonspecific items into specific ones. For example, when respondents report expenditures for meat rather than beef or pork, allocations are made, using proportions derived from item-specific reports in other completed diaries. Income tax data are not estimated for the Diary Survey because the published integrated information uses Interview Survey after-tax income information, and because of the limited amount of background detail collected in the Diary Survey.

Data adjustment methods

Imputation

To publish an accurate estimation of spending for CUs, BLS imputes values for missing or inconsistent data fields. There are four broad types of missing values imputed in the CE: demographic characteristics about the CU and its members; missing values for reported expenditure items; and attributes about a reported expenditure (e.g., whether a car is purchased new or used); and income. Depending on what type of data fields are being imputed, different methods of imputation are used. These methods include:

- Hot deck imputation, where values are copied from other CUs that share similar characteristics;
- Cold deck imputation, where values are copied from households with similar characteristics in another data source;
- Weighted mean imputation, where a weighted average of all values reported by CUs that share similar characteristics is used for the missing value;
- Percent distribution imputation, used for non-numeric attribute information (such as demographic information about members in the CU) where a value is randomly assigned based on the distribution of reported values;
- Regression analysis, where values are predicted using a model of independent variables;
- Multiple imputation, used for income imputation, in which the model is “shocked” with noise to obtain five estimates of income. More details on income imputation can be found in the Data Adjustments section in [CE FAQs](#) and the [user's guide to income imputation in the CE](#); and

- Income tax estimation, where all state and federal income taxes are estimated for all CUs in the Interview Survey, which is used for publication tables. BLS uses an internal version of the National Bureau of Economic Research's TAXSIM software in estimating tax liabilities. Tax liabilities reflect only what TAXSIM estimates is owed, but not necessarily the actual amount that the CU paid. Along with tax liabilities, refundable credits owed to a CU (e.g., additional child tax credit, earned income tax credit) are estimated for each CU regardless of whether the CU received them. For more information about income tax estimation, see the Data Adjustments section in [CE FAQs](#).

Allocation

BLS allocates data to provide information at a sufficient level of detail to meet tabulation requirements. This situation arises when a respondent does not provide the required amount of detail for an expenditure (e.g., the respondent reports "various household appliances" instead of separate reports of a "microwave" and a "blender"). Similar to imputation, BLS uses different methods of allocation dependent on the type of expenditure: distribution ratio, fixed ratio, and probability distribution ratio.

- The distribution ratio is used when specific items within a combined reported item is known and the reported value is assigned in a proportional amount as determined by reported records to those items.
- Fixed ratios are used to assign a proportion of the reported value to specific items based on proportions identified from other data sources.
- The probability distribution ratio is used when specific items in each category are unknown. Percentiles are determined for all potential items in the category and a selected subset of target item codes for which the lower quartile is less than the expenditure reported that requires allocation. A random selection of 6 to 12 targets are chosen, and the mean value for each selected item is subtracted from the combined reported expenditure. This is repeated until the total amount of the reported expenditure is exhausted.

Calculation methodology

After collecting expenditure data from a representative sample of CUs across the nation, the sample is weighted to produce estimates for the U.S. population of CUs as a whole. For this purpose, each CU in the survey is assigned a weight equal to the number of similar CUs in the nation that it represents.

Several factors are involved in computing the weight of each CU from which a usable interview is received. Each CU is initially assigned a base weight equal to the inverse of its probability of being selected for the sample. The probability is the number of addresses selected for the sample divided by the total number of addresses in the Census Bureau's master address file. CE base weights are typically around 10,000, which means a CU in the sample represents 10,000 CUs in the U.S. civilian noninstitutional population—itsself plus 9,999 other CUs that were not selected for the sample. The base weight is then adjusted by the following factors to correct for certain nonsampling errors:

Weighting control factor adjusts for subsampling in the field. Subsampling occurs when a data collector visits a particular address and discovers multiple housing units where only one housing unit was expected.

Noninterview adjustment factor adjusts for interviews that cannot be conducted in occupied housing units because of a CU's refusal to participate in the survey or the inability to contact anyone at the housing unit despite repeated contact attempts. This adjustment is based on region of the country, CU size, number of contact attempts, and the average adjusted gross income in the CU's zip code according to a publicly available database from the Internal Revenue Service.

Calibration factor adjusts the weights to 35 known population counts to account for frame undercoverage. These known population counts are for age, race, household tenure (owner or renter), division of the country, urbanicity (urban or rural), and ethnicity (Hispanic or non-Hispanic). The population counts are updated quarterly using the Current Population Survey (CPS) estimates.¹ Each CU is given its own unique calibration factor. There are infinitely many sets of calibration factors that can make the weights add up to the 35 known population counts. BLS uses nonlinear programming to select the set that minimizes the amount of change made to the "initial weights" (initial weight = base weight x weighting control factor x noninterview adjustment factor).

After adjusting the base weights by these factors, the final weights are typically around 25,000, which means an interviewed CU represents 25,000 CUs in the U.S. civilian noninstitutional population—itsself plus 24,999 other CUs that did not participate in the survey.

Using the weights

Using these weights, the average expenditure per CU on a particular item is estimated with the standard weighted average formula:

$$\bar{y} = \frac{\sum_{i \in S} w_i y_i}{\sum_{i \in S} w_i}$$

where,

\bar{y} = the average expenditure per CU on the item;

y_i = the expenditure made by the i^{th} CU in the sample on the item;

w_i = the weight of the i^{th} CU in the sample; and

S = the set of CUs in the sample that participated in the survey.

For example, if \bar{y} is the expenditure on eggs made by the i^{th} CU in the sample during a given time period, then \bar{y} is an estimate of the average expenditure on eggs made by all CUs in the U.S. civilian noninstitutional population during that period.

Calculation precision

The precision of the estimator \bar{y} is measured by its standard error. Standard errors measure the sampling variability of the CE estimates. That is, standard errors measure the uncertainty in the survey estimates caused by the fact that a random sample of CUs from across the United States is used instead of every CU in the nation. (See table 1.)

The CE's standard errors are estimated by using the method of "balanced repeated replication." In this method, the sampled PSUs are divided into 43 groups (called *strata*), and the CUs within each stratum are randomly divided into two *half samples*. Half of the CUs are assigned to one half sample, and the

other half are assigned to the other half sample. Then 44 different estimates of \bar{y} are created using data from only one half sample per stratum. There are many combinations of half samples that can be used to create these replicate estimates, and the CE uses 44 of them that are created in a “balanced” way with a 44x44 Hadamard matrix. The standard error of \bar{y} is then estimated by:

$$SE(\bar{y}) = \sqrt{\frac{1}{44} \sum_{r=1}^{44} (\bar{y}_r - \bar{y})^2}$$

where \bar{y}_r is the r th replicate estimate of \bar{y} .

The coefficient of variation is a related measure of sampling variability that measures the variability of the survey estimate relative to the mean. It is defined by the equation:

$$CV(\bar{y}) = \frac{SE(\bar{y})}{\bar{y}} \times 100\%$$

Table 1. Precision of the Consumer Expenditure Surveys expenditure estimates, integrated Diary and Interview Survey data, 2019

Item	Average annual expenditure per consumer unit	Standard error, SE(\bar{y})	Coefficient of variation, CV(\bar{y}) (in percent)
Total expenditures	\$63,036	\$578	0.92
Food	8,169	119	1.45
Housing	20,679	195	0.94
Apparel	1,883	69	3.64
Transportation	10,742	194	1.81
Healthcare	5,193	70	1.35
Entertainment	3,090	129	4.19
Personal care	786	14	1.76
Reading	92	6	7.03
Education	1,443	86	5.93

Table 1. Precision of the Consumer Expenditure Surveys expenditure estimates, integrated Diary and Interview Survey data, 2019

Item	Average annual expenditure per consumer unit	Standard error, SE(\bar{y})	Coefficient of variation, CV(\bar{y}) (in percent)
Tobacco products and smoking supplies	320	11	3.57
Miscellaneous	899	43	4.73
Cash contributions	1,995	109	5.48
Personal insurance and pensions	7,165	131	1.83

Source: U.S. Bureau of Labor Statistics.

Integrated survey data

Integrated data from the Interview and Diary Surveys provide an approximately complete accounting of consumer expenditures which neither survey component alone is designed to do. For example, most food expenditures in the integrated data tables come from the Diary Survey, which includes detailed items (e.g., rice, strip steak, breakfast purchased from full-service restaurants) not collected in the Interview Survey, while rent, mortgage interest, utilities, and major appliances, which are less likely to be observed during the week of Diary Survey participation, come from the Interview Survey, for which respondents report based on a three-month recall period.

The Interview Survey collects data on expenditures for overnight travel and information on insurance reimbursements for medical care costs and automobile repairs, which are not collected in the Diary Survey. Based on integrated Interview and Diary Surveys data, expenditure data that come exclusively from the Interview Survey, along with global estimates, such as those for food and alcoholic beverages, average about 95 percent of total estimated spending.

For items unique to one or the other survey, the choice of which survey to use as the source of data is obvious. However, there is considerable overlap in coverage between the surveys. Because of the overlap, the integration of the data presents the problem of determining the appropriate survey component from which to select the expenditure items. When data are available from both survey sources, the more reliable of the two is selected, as determined by statistical methods. The selection of the survey source is evaluated every two years. For more information on the source selection process, see the 2011 Anthology article [CE source selection for publication tables](#).

NOTES

¹ The CPS estimates the total number of people in the U.S. civilian noninstitutional population every month, and BLS averages the three monthly population estimates for its quarterly population estimates. The CPS generates its monthly population estimates by taking the Census Bureau's "official" population estimates from the previous year and updating them to account for the births, deaths, and net migration (immigration minus emigration) that occur each month.

Presentation

Information from the Consumer Expenditure Surveys (CE) is available in [tables](#), [microdata files](#), a [LABSTAT database](#), and publications. The Bureau of Labor Statistics (BLS) also provides outreach with regularly occurring free events like the Survey Methods Symposium and Microdata Users' Workshop. When budget and opportunity permit, BLS staff who work on the CE attend conferences, visit college and university campuses, and arrange meetings with interested parties. For more information, contact the Consumer Expenditure Surveys, Office of Prices and Living Conditions at (202) 691-6900 or by email at cexinfo@bls.gov. CE staff are available Monday through Friday to respond to inquiries. To be notified when new products are available, please sign up for [CE updates](#).

Tables

BLS publishes a standard set of CE tables each year, which includes income quintile, income decile, income class, age of reference person, generation of reference person, selected age ranges of reference person, size of consumer unit (CU), number of earners, composition of CU, Census region of residence, population size of area, housing tenure, race, Hispanic origin, occupation, highest education level of any member, and type of area (i.e., urban or rural). As part of the annual release, the CE program also publishes cross-tabulated tables by age, region, size of CU, and gender; additional detailed geographic breakouts; multiyear tables with means for several years; as well as detailed top line means tables that include the most granular level of expenditure data available, along with variances and percent reporting for each expenditure item. Disaggregated to the most detailed expenditure categories available. Tables going back as far as 1960–61 are available on the CE website. Unpublished, but releasable, tables of detailed expenditures by demographic characteristic can be obtained by sending a request to cexinfo@bls.gov.

Table estimates include some combination of means, shares, and variances. Starting with the 2000 data, estimates of standard error for integrated Diary and Interview Surveys data are available on the BLS website. For more information about the tables and their uses, see the [tables getting started guide](#). BLS also periodically performs special tabulations and makes them available as part of the CE research products page. The current set of research tables includes annual means by income quintiles for selected states, which utilize [state weights](#) and public use microdata ([PUMD](#)). These tables cover 2 years of data to increase the reliability of the data.

Microdata

Microdata for CE are available in two formats—public use microdata (PUMD) and restricted microdata. Both types are available for the Diary and Interview Surveys and contain expenditure and income data for each CU. The PUMD undergo additional processing to protect the identities of respondents by eliminating selected geographic detail and topcoding selected income and expenditure data. Topcoding refers to a confidentiality protection method in which a subset of extremely high or low values are averaged together and the original values are replaced with the average amount. The restricted microdata have not had these steps taken, so more detailed geographic information as well as data not topcoded are available. Restricted data can be accessed by applying to BLS [on-site researcher program](#).

Interview Survey files contain detailed expenditure data in two different formats: MTBI files that present monthly values in an item-coding framework based on the Consumer Price Index (CPI) pricing scheme; and EXPN files that organize expenditures by the section of the Interview Survey questionnaire in which they are collected. Expenditure values on EXPN files cover different time periods depending on the specific questions asked, and the files also contain relevant non-expenditure information not found on the MTBI files. For those interested in examining broader categories of expenditures (e.g., housing or owned dwellings) rather than detailed items (e.g., mortgage interest or property taxes), summary variables are

also available on the FMLI files. The Diary Survey contains detailed expenditure data in the EXPD files, all items in a coding framework based on the CPI, along with family level statistics in the FMLD files.

Users can also use the expenditure data to calculate representative statistics with the weight variable FINLWT21. This variable attributes a weight to each NEWID, the identifying variable for one CU for one quarter, which allows users to estimate values for the entire population. This variable is available in the FMLI and FMLD files.

In addition, CE is researching weights that will allow PUMD to be representative of select states. BLS makes these weights available for use, as well as state level tables, and detailed tables on the [CE research products page](#).

For more information on using the PUMD see the [PUMD getting started guide](#). The annual Interview and Diary Surveys microdata files are available beginning with 1984, as well as for selected earlier years.

LABSTAT

The [CE LABSTAT database](#) provides tools to access historical CE data (1984 onward) to produce trends in expenditures by demographic groups of interest. Documentation on how to use the CE LABSTAT database is available in the [CE LABSTAT getting started guide](#). Not all tabular data are in the CE LABSTAT database. For example, The CE LABSTAT database only contains calendar year data, so midyear tables cannot be replicated using the database. The detail of items in LABSTAT is less than the detail available on standard tables. Similarly, the CE LABSTAT database does not include data on variances or by metropolitan statistical areas. You can find the additional tabulations on the [CE tables](#) page.

Publications

BLS publishes CE [news releases](#) each year along with the calendar year tables. The news releases summarize the estimates and changes associated with each release. Articles that include analyses of CE data are published online in the [Monthly Labor Review](#), [Beyond the Numbers](#), [Spotlight on Statistics](#), in CE data comparisons, and in research reports. For a listing of these and other articles, see the [CE publications](#) page. Other survey information is available on the [CE website](#), including answers to frequently asked questions, copies of the Interview and Diary Surveys, and a glossary of terms for survey products.

CE Survey Methods Symposium and Microdata Users' Workshop

The CE Survey Methods Symposium focuses on survey methodology, and typically features staff who work on the CE and other BLS surveys along with invited researchers who are not affiliated with the BLS. The symposium is typically a 1-day event.

Held over 3 days, the CE Microdata Users' Workshop starts with presentations designed for those who have never used the data and builds to expert topics. The workshop also features presentations from researchers not affiliated with BLS who describe the nature of their projects, specific files and variables they use, the problems (and solutions) they have encountered working with the data, and any other relevant topics they care to share. The workshop also features opportunities to meet one-on-one with a BLS expert who works on the CE to discuss any aspect of a current or potential project, general or specific, about which the attendee has questions or concerns. BLS selects presentations from researchers who answer the call for presenters.

More information about these events is available on the CE Symposium and Workshop website.

History

The Bureau of Labor Statistics (BLS) studies of family living conditions rank among the oldest data collected. The first nationwide expenditure survey was conducted during 1888–91 to study workers' spending patterns as elements of production costs. With special reference to competition in foreign trade, the survey emphasized the worker's role as a producer, rather than as a consumer. In response to rapid price changes prior to the turn of the 20th century, a second survey was administered in 1901. The resulting data provided the weights for an index of prices of food purchased by workers that was used until World War I as a deflator for workers' incomes and expenditures. A third survey, conducted during 1917–19, provided weights for computing a cost-of-living index, now known as the Consumer Price Index (CPI). BLS conducted its next major survey, covering only urban wage earners and clerical workers, during 1934–36, primarily to revise CPI weights.

BLS conducted major survey revisions through the Great Depression and World War II and up through the 1970s. The need for more timely data than could be supplied by surveys conducted every 10 to 12 years—intensified by the rapidly changing economic conditions in the 1970s—led to the initiation of the current continuing survey in late 1979. Unlike the previous surveys, the U.S. Census Bureau, under contract with BLS, conducted all sample selection and field work. Another significant change was the use of two independent surveys to collect the information—a Diary Survey and an Interview Survey. A third major change was the switch from an annual recall to a quarterly recall in the Interview Survey, and daily recordkeeping of expenditures in the Diary Survey. As with the earlier surveys, the resulting data from 1979 onward have been used to revise CPI weights. For details, see the [CPI Handbook of Methods](#). The objectives of the Consumer Expenditures Surveys (CE) remain the same: to provide the basis for revising weights and associated pricing samples for the CPI and to meet the need for timely and detailed information on the spending patterns of different types of families.

Below is a timeline of major events in the history of BLS CE surveys. For a more detailed timeline of the continuous survey since 1979, see [Improvements and protocol changes](#).

Timeline Events:

1888–91: The first nationwide expenditure survey conducted to study workers' spending patterns as elements of production costs.

1901: The second nationwide expenditure survey conducted in response to rapid price changes prior to the turn of the 20th century. It provided the weights for an index of food prices purchased by workers.

1917–19: The third expenditure survey conducted. It provided weights for computing a cost-of-living index, now known as the CPI.

1934–36: Expenditure data collected from only urban wage and clerical workers used in revising CPI weights.

1935–36: The first-ever nationwide rural and urban expenditure survey was collected.

1941–42: Urban and rural expenditure survey conducted during World War II to measure domestic household expenditures during wartime.

1944: The wartime expenditure survey repeated for only urban households.

1950: Expenditure survey conducted for urban households.

1960–61: Expenditure survey for both urban and rural households conducted.

1972–73: First survey collected by the U.S. Census Bureau for BLS. Began the first use of two collection instruments: a weekly Diary Survey and the 3-month recall Interview Survey.

1979: The CE begins continuous monthly data collection in urban and rural areas.

1984: Beginning of annual calendar year published data tables.

1984: Integration of Diary and Interview survey data for publication in CE reports and bulletins.

1992: Published bulletins become reports and free-of-charge to the public.

1995: First release of CE data tables to the public on the CE website.

2000: Release of all historic CE tables for expenditure means from 1984 forward.

2000: First release of standard error tables on the CE website.

2002: CPI begins using biennial CE weights to update CPI cost weights every 2 years instead of updating the CPI every 10 years using 3 years of CE expenditures.

2003: Computer assisted personal interview (CAPI) starts.
2003: First release of CE anthology publication.
2004: Introduction of imputed income to fill in all missing income values.
2005: Introduction of a more user-friendly Diary Survey.
2009: The Gemini Project CE Redesign long-term research begins. The primary mission of the Gemini Project is to improve data quality through a verifiable reduction in measurement error, with a particular focus on underreporting.
2013: First publication of midyear data tables every 12 months.
2013: Introduction of estimated federal and state income taxes with the published 2013 data tables. Replaced all collected and missing amounts with estimated amounts.
2015: Noninterview adjustment calculations include income as a weighting variable.
2015: Initial "bounding" Interview Survey dropped. Number of interviews per household drops from five to four.
2016: Publication of public use microdata (PUMD) from 1996 forward.
2017: First release of state-level weights, for use with 2016 data.
2020: Publication of previously unavailable PUMD on the CE website for 1980–96.
2021: First release of midyear PUMD, to provide data at the earliest time possible on how the COVID-19 pandemic affected consumer expenditures.
2022: Official addition of an online mode for the Diary Survey.

More information

Further information on the Consumer Expenditures Surveys (CE) program can be found through the CE website, at <https://www.bls.gov/cex>.

Survey methods research

CE undergoes continuous evaluation, by comparing results with other sources and by performing internal statistical, qualitative, and cognitive analyses to address current methodological concerns. To improve expenditure estimates, in the mid-1980s the Bureau of Labor Statistics (BLS) began CE research that was related to the data collection instruments, field procedures, and sources of potential survey error, and has since become standard practice. In 1999, BLS established a separate Branch of Research and Program Development (BRPD) within the Division of Consumer Expenditure Surveys, with the mission to improve CE survey data collection procedures, data quality, and cost efficiencies through the development, implementation, and analysis of methodological studies and research projects. In recent years, BRPD has focused on four core areas: the Gemini Project to redesign the survey; analyzing historical data in support of ongoing methodological improvements; field testing different data collection methods; and exploring the use of alternative data in the context of the CE.

The Gemini Project

BLS began the Gemini Project in 2009 with a goal of redesigning the CE. Named for the two component surveys in the CE (Interview Survey and Diary Survey), the Gemini Project was created in response to increasing evidence of measurement error, declining response rates, the emergence of new data collection technologies, and the need for more flexibility in addressing changes in the interviewing environment. The primary mission of the Gemini Project is to improve data quality through a verifiable reduction in measurement error, with a particular focus on underreporting. Early stages of the project focused on gathering facts to inform redesign decisions. This included conducting and reviewing research on survey methodologies and prioritizing user needs.

Additionally, in 2010, BLS contracted with the Committee on National Statistics (CNSTAT) to convene an expert panel charged with recommending different CE design options that would meet the project goals. The CNSTAT panel presented three alternate designs in September 2012. In 2013, the CE program

approved a comprehensive redesign proposal based on 3 years of information gathering, inquiry, and synthesis, including a review of the CNSTAT recommendations. The redesign proposal meets key stakeholder requirements and addresses three factors believed to affect the survey's ability to collect high quality data; specifically, measurement error, environmental changes, and flexibility. Since the release of the redesign plan, BRPD has field tested different components of the plan for implementation viability. These tests included a [proof-of-concept test](#), an [incentives test](#), an [individual diaries test](#), and an [online diary field test](#). A phased implementation plan was made based on findings from these tests. The phased implementation began with an optional online diary mode for the CE Diary Survey in 2022 and will continue with the introduction of a streamlined questionnaire in the CE Interview Survey in 2023. For further information on the Gemini Project, including information about current research studies and the project's timeline, see the [Gemini Project webpage](#).

Research overview

BRPD conducts ongoing research, both in support of the redesign effort and to continuously improve data quality and data collection procedures while balancing survey costs.

Current research has focused on analyzing historical data in support of methodological improvements, field testing alternate protocols, and investigating sources of alternative data. The first area of research is useful for reviewing the existing survey protocols and considering the potential impact of design changes. The second area of research provides empirical insight for decisions on implementing future protocol improvements. Finally, the last area of research allows us to identify potential sources that could be used to complement or replace components of the surveys. More information on the BLS investigation of alternative data can be found in the February 2021 *Monthly Labor Review* article "[A framework for the evaluation and use of alternative data in the Consumer Expenditure Surveys](#)." Details about findings from recently completed research projects are provided in the [CE library](#).

Other ongoing survey improvements

In a collaborative effort headed by the CE branch of Production and Control involving the different CE branches and divisions, there are regular biennial Interview Survey questionnaire revisions and other improvements. These improvements include adding new products into the survey, deleting outdated wording or categories, improving noninterview adjustment through the inclusion of income data at the zip code level, using TAXSIM to provide estimated income taxes, and publishing new tables.

Last Modified Date: September 12, 2022