The Consumer Expenditure Survey Data Quality Profile
(Prototype Iteration 2)

August 16, 2017

Evan Hubener, Clayton Knappenberger, Julie Sullivan, and Lucilla Tan

CE Surveys Program Report Series

Overview

The Consumer Expenditure Survey (CE) has historically provided some limited metrics for data users to evaluate the overall quality of output provided in its products. Published tables provide standard errors; the public-use microdata user guide provides response rates, and the datasets contained in the public-use microdata provide all the variables and flags necessary for users to create his or her own quality measures. There has long been a recognition for the need for more comprehensive data quality metrics that are timely, routinely updated, and accessible to data users from a single source, a Data Quality Profile (DQP). However, there is also recognition of the high cost in terms of resources and commitment to identifying appropriate metrics and establishing the information base necessary to routinely produce reports on survey data quality. In order for this effort to be sustainable, the benefits from it must be relevant and useful to survey operations and data users.

This report, the CE Data Quality Profile Prototype version 2 (DQP2), is the *second* in a series of iterations towards developing a single reference on a comprehensive set of CE data quality metrics that are timely and routinely updated for the Consumer Expenditure Interview Survey (CEQ) and the Consumer Expenditure Diary Survey (CED). Recognizing the benefits of "learning-by-doing" – using cumulative experience to provide CE with a practical understanding of what resources are needed and how best to deploy them to routinely produce a DQP - *the first* iteration DQP1 produced a small set of metrics with limited resources. The goals of this second iteration were to refine and expand the set of metrics reported in DQP2, and begin to identify the resources needed to build the infrastructure to produce these metrics routinely.

Content links

Visual Summary

Metrics:

- Response rates: official published tables
- Use of Records in the CEQ
- Expenditures edit rate: processed data
- Income Imputation Rates

Appendix: detailed definitions

- Response and nonresponse rates
- Records use comparison groups
- Survey mode comparison groups
- Processed expenditures, Reported item counts
- Expenditure edit type
- Income edit type

The metrics described in this report are for the years 2010 through 2015. The associated data quality dimension(s) of each metric is summarized in the table below:

Metric	Total Survey Error dimensions associated with the metric		
	Measurement	Nonresponse	Processing
Final Disposition Rates of eligible units, CED & CEQ (official tables/SMD) • unweighted and base-weighted		★	
Records Use by CEQ respondents • conditioned on mode • reported number of items and total expenditures	★		
CEQ Processed Expenditure Data Edit Rate • conditioned on survey mode	★	★	★
CED & CEQ Processed Income Data Edit Rate	★	★	★

In the next section, we present visualizations that highlight findings about the metrics. Detailed metric tables and descriptions then follow, and finally, definitions used to construct the metrics appear in the Appendix.

Visual Summary

In this section, metric trends for the reporting period (2010 through 2015) are highlighted in a panel of graphs (Figure 1). Further details about the individual metrics and detailed data tables are in the sections that follow the visual summary.

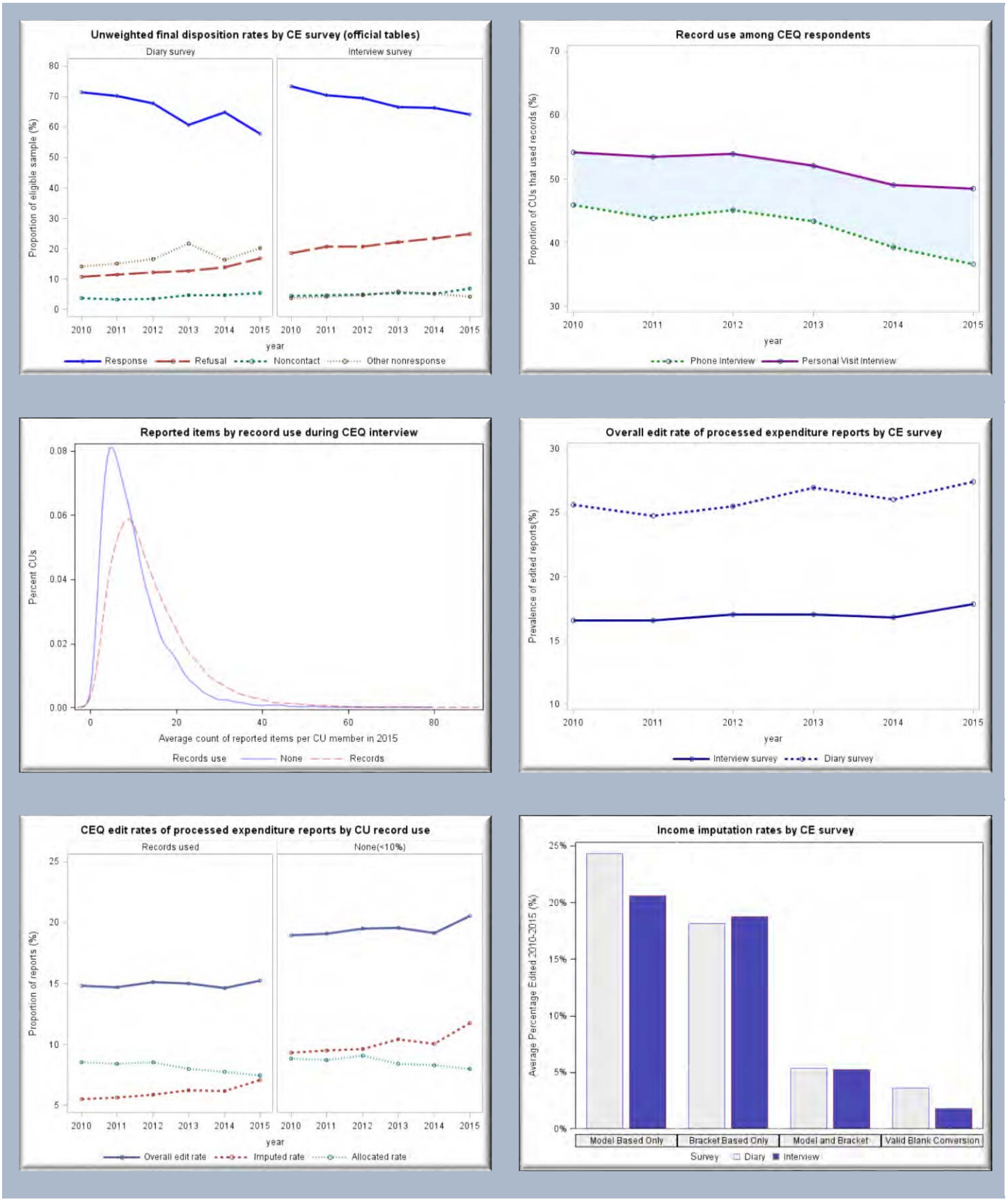
Metrics with declining trends

- Unweighted response rates for the universe of CUs whose data are used to produce CE's official tables continued to decline for both CE surveys in 2015 (Figure 1 , top of left panel). Base-weighted rates were similar to unweighted rates. The trend of declining response rates call for understanding if and how differences between respondents and nonrespondents affect nonresponse bias of key survey estimates.
- The prevalence of overall record use for the CEQ declined across the reporting period. Record use was higher among personal visit than phone survey modes, but declined for both modes (Figure 1 , top of right panel). This declining trend has an adverse impact on measurement error; among respondents who use records, the average number of reported items per CU member is higher (Figure 1 , middle left panel), as are total expenditures. Also, for these respondents, the overall edit rate of processed expenditure reports is lower (Figure 1 , bottom of right panel).

Metrics with rising trends

- Survey refusal and noncontact rates were higher in the second half of the reporting period than the first for both the CEQ and the CED (Figure 1 , top of left panel). Other nonresponse rates (reasons for non-response other than refusal and noncontact) are higher in the CED than in the CEQ throughout the reporting period; this difference can be partially explained by diaries classified as nonresponse due to "Diaries Placed Too Late " – a nonresponse code that does not exist in CEQ. Beginning in 2017, interviewers will have the full month instead of just 7 days to place a Week 1 diary. We therefore expect the difference between CED and CEQ other nonresponse rates to be reduced in 2017.
- Overall edit rates of processed expenditure reports are higher in the CED than the CEQ (Figure 1 , top right panel). The gradual increase in the overall edit rate of these reports was driven by the rise in the imputation rate for the CEQ in 2015. For the CED, allocation is the dominant expenditure edit method, and the allocation rate also increased in 2015. Since higher imputation rates imply higher rates of item nonresponse, and higher allocation rates imply incomplete responses (insufficient detail), an increase in these rates adversely affects measurement error. Increasing edit rates also increase the risk of processing error. If this trend continues, an attempt to understand if there are sections of the survey that are driving the higher edit rates would inform possible design interventions.
- The trend for the prevalence of *no* imputed income sources in "Total CU income before tax" continued to rise for both the CEQ and the CED in 2015, due to declining model-based imputation rates. Model-based imputation rates have been the dominant income imputation method (Figure 1 , bottom of right panel). While this is a desirable trend for measurement error, more than a third of CUs continue to have some income component of total income that is imputed. The CE will continue to investigate and monitor how imputation rates vary among the income components.

Figure 1. Select metric trends from 2010 to 2015



1. Final Disposition Rates

The unit of observation for the CE is the Consumer Unit (CU), so response and nonresponse rates are computed at the CU level. The CE adopts the Census Bureau's categorization of eligible CUs who do not respond to the survey as a "Type A nonresponse". Type A nonresponse is differentiated into subcategories of reasons for nonresponse: "*Noncontact*" when the interviewer is unable to contact an eligible member of the CU; "*Refusal*" when the contacted CU member refuses; and "*Other nonresponse*" for miscellaneous other reasons. Among the "Type A Other nonresponse" reasons is a minimal expenditure edit check performed at BLS (CE) that could change an interviewer-coded "completed interview" for a CU to "nonrespondent"; this type of edit is referred to as the *nonresponse reclassification*. A mapping of the CE to AAPOR final disposition codes for in-person, household survey is presented in the Appendix.

Response and nonresponse rates are measures of cooperation levels in a survey. Since not all eligible sample units will be available or agree to participate in the survey, there will be some nonresponse to the survey request. Characteristics of nonrespondents may differ from respondents, and if these characteristics correlate with their expenditures, their omission from the survey may result in bias in the estimates produced from the survey. While weighting adjustments may reduce bias, the effectiveness of this approach depends on the availability and quality of variables used in the weighting and so concerns about bias persist. A single, survey-level measure, such as a survey response rate, in itself is an inadequate measure of nonresponse error. Nevertheless, higher response rates are preferred in the absence of other indicators of nonresponse bias.

The nonresponse reclassification is conducted in both the Interview and Diary surveys. The nonresponse reclassification rates can serve as an indicator of the potential for nonresponse bias because the minimal expenditure edit (which triggers reclassification) converts these 'respondents' to nonrespondents. If those reclassified as nonrespondents are systematically different from respondents, nonresponse bias will result. Thus, ceteris paribus, lower reclassification rates are desired.

Response rates are reported unweighted and weighted. *Unweighted response rates* provide an indication of the proportion of the sample that resulted in useable information to produce estimates, They also serve as a useful means of monitoring the progress of fieldwork and for identifying problems with nonresponse that can be addressed during fieldwork operations. *Weighted response rates* provide an indication of the proportion of the survey population for which useable information is available, since the weights allow for inference of the sample to the population. The weights used are base weights (the inverse probability of selecting the sample units). Both rates are usually similar unless the probabilities of selection and the unit response rates in the categories with different selection probabilities vary considerably.

Each survey wave is treated independently (that is, as an independent CU) in the computation of the CE Interview Survey response rates, and each diary week is treated independently for the CE Diary Survey rates. So the rates reported in this section are <u>not</u> longitudinal response rates.

a. Official published tables

CE Interview Survey (CEQ)

Unweighted rates: official CE tables. CEQ *response rates* steadily declined from 73.4 percent to 64.2 percent between in 2010 and 2015 (Table 1.1). The drop of 3 percentage points in 2013 could be partially accounted for by the Federal Government shut down that affected data collection in October, but another dip of 2 percentage points occurred in 2015. This may be partly attributable to a change in data collection staff arising from a sample redesign - there is a loss of some experienced CE data collectors in areas dropped, and the addition of new, less experienced CE data collectors in the areas added.

Refusal rates rose from 18.6 percent to 21.2 percent between 2010 and 2015, with a smaller increase in *unweighted noncontact rates* from 4.3 percent to 6.8 percent over the same period. The *nonresponse reclassification rate* declined from 2.1 percent to 0.8 percent of "Other nonresponse" cases between 2010 and 2015 (Table 1.3).

Base-weighted rates: official CE tables. The magnitudes and trends of the CEQ *base-weighted* response rates, refusal rates, and noncontact rates were similar to their respective unweighted rates (Table 1.2).

Table 1.1 CEQ Distribution of Final Dispositions for Eligible CUs (official tables, unweighted)

Year*	No. Eligible CUs	Interview	Refusal	Noncontact	Other Nonresponse
	Row Percent Distribution				
2010	38,718	73.4	18.6	4.3	3.7
2011	38,348	70.4	20.8	4.6	4.2
2012	38,835	69.5	20.8	5.0	4.7
2013	39,142	66.7	22.1	5.4	5.8
2014	39,003	66.4	23.3	5.2	5.0
2015	36,692	64.2	24.8	6.8	4.2

* Excludes bounding interviews prior to 2015, and excludes Jan 2015 data due to sample redesign.

Table 1.2 CEQ Distribution of Final Dispositions for Eligible CUs (official tables, base-weighted)

Year*	Interview	Refusal	Noncontact	Other Nonresponse
	Row Percent Distribution			
2010	73.8	18.2	4.3	3.7
2011	70.4	20.7	4.6	4.3
2012	69.7	20.8	4.8	4.8
2013	66.8	22.1	5.3	5.8
2014	66.6	23.2	5.1	5.1
2015	63.8	25.1	6.9	4.2

* Excludes bounding interviews prior to 2015, and excludes Jan 2015 data due to sample redesign.

Table 1.3. CEQ: Prevalence of nonresponse reclassifications (official tables, unweighted)

			Nonresponse Reclassifications		
Year*	No. Eligible CUs	No. Other Nonresponse	No. CUs	Other Nonresponse (%)	Eligible CUs (%)
2010	38,718	1,427	30	2.1	0.077
2011	38,348	1,606	24	1.5	0.063
2012	38,835	1,816	13	0.7	0.033
2013	39,142	2,258	18	0.8	0.046
2014	39,003	1,960	10	0.5	0.026
2015	36,692	1,537	13	0.8	0.035

* Excludes bounding interviews prior to 2015, and excludes Jan 2015 data due to sample redesign.

CE Diary Survey (CED)

Unweighted rates: official CE tables. While CED response rates were consistently lower than the CEQ between 2010 and 2015, they mirrored the same declining trend of the CEQ. The CED *response rate* fell from 71.5 percent in 2010 to 57.7 percent in 2015 (Table 1.4), with the largest annual percentage points drop of 7 percent in 2013 and 2015. The drop off in the response rate in 2013 can partially be explained by the shutdown of the Federal Government, and partially in 2015 by an increase of 3 percentage points in the CED refusal rate. The CED *refusal rate* has trended upwards but were consistently lower than those for the CEQ, rising from 10.8 percent in 2010 to 16.7 percent in 2015. The CED *noncontact rate* increased from 3.6 percent to 5.4 percent during the same period.

The nonresponse reclassification rate as a proportion of other-nonresponse is higher in the CED than the CEQ, but it had declined from 43.0 percent to 25.2 percent of other non-response cases between 2010 and 2015 (Table 1.6). Other nonresponse rates (reasons for non-response other than refusal and noncontact) are higher in the CED than in the CEQ throughout the reporting period; this difference can be partially explained by diaries classified as nonresponse due to "Diaries Placed Too Late " – a nonresponse code that does not exist in CEQ. Beginning in 2017, interviewers will have the full month instead of just 7 days to place a Week 1 diary. We therefore expect the difference between CED and CEQ other nonresponse rates to be reduced in 2017.

Base-weighted rates: official CE tables. The magnitudes and trends of the CED *base-weighted* response rates, refusal rates, and noncontact rates were similar to their respective unweighted rates (Table 1.5).

Table 1.4 CED Distribution of Final Dispositions for Eligible CUs (official tables, unweighted)

Year	No. Eligible CUs	Interview	Refusal	Noncontact	Other Nonresponse
		Row Percent Distribution			
2010	19,988	71.5	10.8	3.6	14.1
2011	19,823	70.2	11.4	3.3	15.1
2012	20,298	67.8	12.1	3.5	16.6
2013	20,296	60.8	12.8	4.7	21.7
2014	20,476	65.0	13.9	4.7	16.4
2015	20,517	57.7	16.7	5.4	20.2

Table 1.5 CED Distribution of Final Dispositions for Eligible CUs (official tables, base-weighted)

Year	Interview	Refusal	Noncontact	Other Nonresponse
	Row Percent Distribution			
2010	71.9	10.7	3.4	14.0
2011	70.3	11.4	3.1	15.2
2012	67.7	12.2	3.3	16.7
2013	60.7	12.8	4.6	22.0
2014	64.8	14.0	4.7	16.5
2015	57.7	16.9	5.4	19.9

Table 1.6. CED: Prevalence of nonresponse reclassifications (official tables, unweighted): 2010-2015

			Nonresponse Reclassifications		
Year*	No. Eligible CUs	No. Other Nonresponse	No. CUs	Proportion of Other Nonresponse (%)	Proportion of Eligible CUs (%)
2010	19,988	2,811	1,209	43.0	6.0
2011	19,823	3,000	1,129	37.6	5.7
2012	20,298	3,370	1,109	32.9	5.5
2013	20,296	4,411	1,112	25.2	5.5
2014	20,476	3,357	1,141	34.0	5.6
2015	20,517	4,141	1,045	25.2	5.1

2. Use of Records in the CEQ

Responses based on expenditure records to survey questions about spending result in higher reporting accuracy and lower measurement error. Thus, a higher prevalence of record use is desirable.

Use of records during the interview

The creation of the comparison groups for records usage analysis is based on the overall *records used* question asked of the interviewer at the end of the CEQ survey (for details, see Appendix). It is optional for respondents to use records, so it is likely that respondents who do choose to use *any* records and refer to them on an as needed basis are more engaged than those respondents who choose not to consult records. In addition, it is plausible that "no or very few records were used" would be more salient in the interviewer's recollection of the interview than the varying extent of records used in the other response options. For these reasons, and for simplicity of interpretation, two comparison groups were created for analysis: "*Records*" vs "*None*", where the None group consisted of CUs whom the interviewer reported as using records "never or almost never (less than 10% of the time)". These two groups comprised roughly half the respondents in 2010. (Table 2.1). However, the prevalence of no records used during an interview increased by 6.8 percent between 2010 and 2015.

Survey mode

Three comparison groups for survey mode were created for analysis: the "*Personal Visit*", "*Phone", and "Mixed" mode* groups (for details, see Appendix). Most interviews were completed using a single mode, with less than 1.5 percent by Mixed mode between 2010 through 2015 (Table 2.1). The dominant survey mode is by Personal Visit, comprising about two-thirds of the completed interviews. After about a 3 percentage points decline in Personal Visit interviews in 2013 and 2014 from 2012, there was an upturn in 2015 (66.7 percent Personal Visit interviews). This may likely be due to the CEQ survey panel changing from five to four waves. Prior to 2015, 1st interviews were 20% of the sample, thereafter, they account for 25% of the sample. Sample units are more likely to be interviewed in person on their first survey wave.

Records use by survey mode

Consistent with expectations that face-to-face interactions provide more opportunities to encourage record use, the prevalence of record use was higher among Visit than Phone interviews between 2010 and 2015 (Table 2.2). However, the trend of records use among Visit interviews had steadily declined from 54.2 percent in 2010 to 48.5 percent in 2015, a fall of 5.7 percentage points. A steeper rate of decline was observed among Phone interviews over the same period, a fall of 9.2 percentage points. It is important to point out that while the 'Mixed' mode has the lowest prevalence of respondents that used no records, these respondents comprised only about 1 percent of all households in the years 2010-2015 (Table 2.1).

CEQ Reported Number of items and Expenditures by Records Use

CE's ultimate goal in encouraging in person interviews and respondent record use is higher quality data. Our hope is that record use would first, help a respondent remember to report items they otherwise would not have reported, and second to remember the correct cost of a certain item purchased. Therefore we relate records use to the number of unique expenditures reported and to the overall amount reported. To control for the size of the CU, both measurements are divided by the number of people that are in the CU. The number of expenditure items reported by the CU was computed from the *reported expenditure records* file (see Appendix for details about how this file was created).

The average and median number of item reports per CU member was about 4 additional reports among the Records group compared with the None (no records) group (Table 2.3). The similarity in the difference between the averages and the medians between the Records and the None groups suggests that this result was not skewed by outliers. This finding confirms our belief that encouraging records use prompts the recollection of items that would otherwise have gone unreported.

Mean expenditures per person for the Records group were consistently higher than the None (no records) group between 2010 and 2015; in 2015, the difference in means between the two groups was $1,498 (Table 2.4). However, the difference in the medians was $1,275 suggesting that one of the means was more skewed by a heavy right tail distribution. Our investigation into CE's expenditure edit rates showed that the imputation rate among respondents who never used records was 5 percentage points higher than for respondents who used records (Table 3.3). Therefore, the disparity in the above means is reduced because of the higher imputation rate for respondents that did not use records.

Table 2.1: Prevalence of records use and survey mode among CEQ respondents

		Survey Mode				Record Use Group		
		Missing	Mixed	Phone	Personal Visit	Missing	None	Records
Year	No. CUs	Row Percent Distribution				Row Percent Distribution		
2010	28,429	0.3	0.9	34.0	64.8	0.7	48.0	51.3
2011	26,990	0.3	1.0	34.7	64.0	0.9	49.1	50.0
2012	26,993	0.3	0.9	33.6	65.2	0.8	48.5	50.7
2013	26,108	0.3	1.1	35.9	62.8	0.8	50.4	48.8
2014	25,908	0.3	0.9	36.5	62.2	0.7	54.0	45.3
2015	23,574	0.4	1.2	31.8	66.7	0.7	54.8	44.5

Table 2.2: Prevalence of records use by survey mode in the CE Interview

Survey Mode	Mixed			Phone			Personal Visit		
Records Use	Missing	None	Records	Missing	None	Records	Missing	None	Records
	Row Percent Distribution			Row Percent Distribution			Row Percent Distribution		
2010	2.2	39.3	58.5	0.7	53.4	45.9	0.3	45.5	54.2
2011	1.8	37.5	60.7	0.7	55.5	43.8	0.4	46.1	53.5
2012	3.6	44.6	51.8	0.6	54.4	45.1	0.4	45.8	53.9
2013	1.1	46.5	52.5	0.6	55.9	43.4	0.4	47.6	52.1
2014	1.7	51.9	46.4	0.4	60.3	39.3	0.3	50.6	49.1
2015	0.7	55.3	44.0	0.3	62.9	36.7	0.2	51.3	48.5

Table 2.3: CEQ respondents: distribution of counts of reported items per CU member by records use

Records Use	**None**				**Records**			
Year	P25	Mean	Median	P75	P25	Mean	Median	P75
2010	5.0	10.2	8.5	13.5	7.5	14.3	12.3	19.0
2011	5.0	10.4	8.7	13.7	7.7	14.6	12.5	19.0
2012	5.0	10.2	8.5	13.3	7.5	14.3	12.0	19.0
2013	5.0	9.9	8.0	13.0	7.3	14.3	12.0	19.0
2014	5.0	9.9	8.0	13.0	7.5	14.4	12.0	19.0
2015	5.0	10.0	8.0	13.0	7.5	14.1	12.0	18.0

Table 2.4: CEQ respondents: distribution of reported expenditures ($) per CU member by records use

Records Use	**None**				**Records**			
Year	P25	Mean	Median	P75	P25	Mean	Median	P75
2010	2,220	4,716	3,519	5,674	3,011	6,269	4,781	7,643
2011	2,264	4,872	3,655	5,954	3,043	6,335	4,868	7,693
2012	2,347	5,118	3,763	6,099	3,150	6,534	5,019	7,892
2013	2,427	5,080	3,833	6,143	3,227	6,719	5,095	8,120
2014	2,486	5,318	3,960	6,438	3,367	6,870	5,306	8,284
2015	2,579	5,646	4,134	6,804	3,448	7,144	5,409	8,581

3. Expenditure data edit rate

At the completion of an interview, data from the interviewer's laptop are transmitted to the Census Master Control System. The Census Bureau's Demographics Surveys Division preforms some preliminary processing and reformatting of the data before transmitting the data to BLS on a monthly basis. At BLS, a series of automated and manual edits are applied to the data in order to ensure consistency, fill in missing information, and to correct errors in the collected data. (For more description about the data collection and processing for the CE surveys, see Handbook of Methods: Consumer Expenditure Survey).

Edits are defined as any changes in the data made during processing that require judgment or assumptions (whether model based or manually adjusted by an economist). Imputation and allocation are two major types of data edits to improve estimates derived from the Interview and Diary Surveys (see Appendix for definitions of edits):

- *Data imputation* edits account for missing or invalid entries and currently apply to most expenditure and income fields, excluding assets.
- *Allocation* edits are applied when respondents provide insufficient detail to meet tabulation requirements. For example, if a respondent provides a non-itemized overall expenditure report for the category of fuels and utilities, that overall amount will be allocated to the target items mentioned by the respondent (such as natural gas and electricity).

In addition to allocation and imputation, data are reviewed and manually edited as needed by BLS economists based on their research and expert judgment.

The need for data imputation results from missing data (item or price nonresponse). Thus, lower imputation rates are desirable. The need for data allocation is a consequence of responses that did not contain the required details of the item asked by the survey. Likewise, lower allocation rates are also preferred, and in general, lower data editing rates are preferred since that lowers the risk of processing error. However, imputation based on sound methodology can improve the completeness of the data and improve overall survey estimates.

Processed expenditure reports

In this section, we describe the edit rates of *processed* expenditure data. Processed expenditure data refer to the final set of expenditure records at the end of data editing, which includes new records that are the result of an edit (for example, an original single report of "groceries" may result in 4 additional records after allocation). Processed expenditure data are used in the CE microdata files used by external data users, as well as the data used to produce estimates that appear in the official CE tables. For the CEQ, we also describe the edit rates conditioned on the respondents' use of records during the interview. The definition of the records use comparison groups, "Records" and "None, is described in the Appendix.

CE Interview Survey (CEQ)

The increase in the overall proportion of edited expenditure reports between 2010 and 2015 was small (from 16.6 percent to 17.9 percent, respectively), with a larger increase in imputation rates from 7.1 percent to 9.4 percent over the same period (Table 3.1).

Use of records during the interview resulted in a higher number of expenditure reports (Table 3.2; and also seen in Table 2.3). In addition, the prevalence of unedited expenditure reports is higher, with imputation rates and allocation rates lower among record users each year between 2010 and 2015, but the difference in allocation rates between the 2 groups was less than 1 percentage point (Table 3.3). The Records group had an average rate of unedited expenditure reports of 85.1 percent compared to 80.7 percent for respondents who did not use records (average of unedited rates over the 6 years for each group in Table 3.3). Average imputation rates for the six years was 6.0 percent among respondents who used records compared to 10.0 percent among respondents who did not use records.

CE Diary Survey (CED)

The CED has a higher edit rate for expenditure reports compared to the CEQ, but like the CEQ, the increase in the overall proportion of edited CED expenditure reports between 2010 and 2015 was small (from 25.6 percent to 27.4 percent, respectively; Table 3.4). Due to the nature of the diary survey, where it is not possible to comprehensively prompt the diary keeper for item detail or categorization required, allocation is the primary edit and generally occurs where multiple items are purchased but one total cost is reported for those items. Between 2010 and 2015, the average rate of unedited expenditure records was 74.0 percent, and 25.8 percent of records are allocated (average of the 6 years of rates in Table 3.4).

Table 3.1 CEQ processed expenditure reports: edit rates

		Type of Edit				
		Allocated*	Imputed & Allocated	Imputed	Other Edit	Unedited
Year	No. Expn Reports	Row Percent Distribution				
2010	2,272,431	8.7	0.1	7.1	0.7	83.4
2011	2,178,837	8.5	0.1	7.3	0.6	83.4
2012	2,192,794	8.8	0.1	7.5	0.6	83.0
2013	1,968,739	8.2	0.1	8.1	0.6	83.0
2014	1,901,766	8.0	0.1	8.0	0.6	83.2
2015	1,752,344	7.7	0.2	9.4	0.6	82.1

* One .G record included in "Allocated" count.

Table 3.2. CEQ: No. of processed expenditure reports by CU's records use status

Year	Records Use Group			
	Missing	None	Records Used	
	No. of Reports			Total
2010	16,327	946,431	1,309,673	2,272,431
2011	16,908	932,798	1,229,131	2,178,837
2012	15,478	928,292	1,249,024	2,192,794
2013	12,744	863,899	1,092,096	1,968,739
2014	11,679	909,562	980,525	1,901,766
2015	11,578	862,003	878,763	1,752,344

Table 3.3 CEQ processed expenditure records: edit type rate by Records use

Year	No. Expn Reports	Other	Allocated*	Imputed & Allocated	Imputed	Other Edit
		Row Percent Distribution				
Records Used						
2010	1,309,673	0.7	8.6	0.0	5.5	85.2
2011	1,229,131	0.6	8.4	0.1	5.6	85.3
2012	1,249,024	0.7	8.5	0.1	5.9	84.9
2013	1,092,096	0.7	8.0	0.0	6.2	85.0
2014	980,525	0.7	7.7	0.1	6.1	85.3
2015	878,763	0.6	7.4	0.1	7.1	84.8
None						
2010	946,431	0.6	8.8	0.2	9.2	81.2
2011	932,798	0.6	8.7	0.2	9.4	81.1
2012	928,292	0.6	9.1	0.2	9.5	80.6
2013	863,899	0.5	8.4	0.2	10.3	80.5
2014	909,562	0.6	8.3	0.2	10.0	81.0
2015	862,003	0.5	8.0	0.2	11.7	79.5
Missing						
2010	16,327	0.8	8.8	0.8	17.6	72.0
2011	16,908	0.8	8.6	0.6	16.3	73.8
2012	15,478	0.8	8.5	0.8	14.7	75.1
2013	12,744	0.5	8.3	0.5	16.9	73.8
2014	11,679	0.8	8.4	0.4	16.8	73.6
2015	11,578	0.8	5.8	0.8	15.0	77.6

* One .G record included in "Allocated" count.

Table 3.4 CED processed expenditure records: edit rate

		Type of Edit		
		Allocated *	Other Edit	Unedited
Year	No. Expn Reports	Row Percent Distribution		
2010	499,712	25.4	0.2	74.4
2011	494,069	24.6	0.2	75.2
2012	487,101	25.3	0.2	74.5
2013	432,184	26.8	0.2	73.0
2014	458,421	25.9	0.1	74.0
2015	412,635	27.4	0.1	72.6

*It is possible for a record to have been split into multiple records by allocation and the allocated records manually corrected to a single record without the allocation variable being reset to 0.

4. CE Income Imputation Rates

CE performs three edits relating to the imputation of income for both the Diary and Interview Surveys. The first is referred to as "Model-based" imputation and occurs when a CU indicates an income source but fails to report the amount of income received. The second is referred to as "Bracket response" imputation, and it occurs when a CU indicates the receipt of an income source and fails to report the amount of income, but does provide a bracket range estimate of the amount of income received. The third income edit is referred to as "All valid blank" conversion, and it occurs when a CU reports that they received no income from any source, but the BLS determines that some income was received from at least one source. For income imputation flag values, see Appendix. Since the need for imputation reflects item nonresponse or insufficient item detail was provided, lower imputation rates are desirable for lowering measurement error. However, imputation based on sound methodology can improve the completeness of the data.

CE Interview Survey (CEQ)

The proportion of CUs whose total income before tax included *no imputed income sources* followed an increasing trend from 52.5 percent in 2010 to 56.4 percent in 2015 (Table 4.1). This increasing trend appears to be driven by declining rates of model-based and bracket response imputation which are down by 2.1 and 0.9 percentage points from 2010 to 2015. The proportion of CUs whose total income before tax included both model-based and bracket response imputations has also decreased over this period by 0.8 percentage points. Overall, the proportion of CUs converted from zero income to a positive income amount remained stable from 2010 to 2015 despite slight increases to 2.0 percent in 2013 and 2.1 percent in 2014.

CE Diary Survey (CED)

As with the CEQ, the proportion of CUs whose total income before tax included no imputed income sources continued an increasing trend from 46.3 percent in 2010 to 50.1 percent in 2015 (Table 4.2). This increasing trend appears to be driven by declining rates of model-based and bracket response imputation which are down by 1.9 and 1.7 percentage points from 2010 to 2015. The proportion of CUs whose total income before tax included both model-based and bracket response imputations has also decreased over this period by 1.0 percentage points. However, the proportion of CUs converted from zero income to a positive income amount has increased from 3.3 percent to 4.1 percent.

NOTE: Drop offs in income imputation rates occurred for both surveys. In 2013, substantial revisions were made to the income section: some questions were merged together or split into new questions. The change in questionnaire is likely the cause of the decrease in income imputation rates for both CED and CEQ in 2013. However, the continuing decline in income imputation rates observed in 2014 and 2015 are likely not attributable

to this questionnaire change. If this trend continues, additional research will be necessary to determine what is driving this decline in income imputation rates, especially if expenditure edit rates continue to rise. From a data quality perspective, a decline in income imputation rates at first seems to be an unequivicably positive trend. A declining imputation rate suggests that respondents are providing higher quality income reports, removing the need to perform edits. However, declining income imputation rates could also indicate an overall decline in income reports if respondents are simply not reporting income sources. The rise in valid blank conversion imputation coupled with the increasing rates of expenditure imputation suggest that this question deserves additional investigation.

Table 4.1: CEQ income imputation rates

Year	No. CUs	Not Imputed	Model Imputation Only	Bracket Imputation Only	Model and Bracket Imputation	Valid Blank Converted
		Row percent distribution				
2010	28,429	52.5	20.8	19.5	5.6	1.6
2011	26,990	52.5	20.4	19.8	5.7	1.6
2012	26,993	52.3	21.1	19.6	5.4	1.5
2013	26,108	53.8	21.4	17.6	5.2	2.0
2014	25,908	54.4	21.1	17.7	4.7	2.1
2015	23,574	56.4	18.7	18.6	4.8	1.6

Table 4.2: CED income imputation rates

Year	No. CUs	Not Imputed	Model Imputation Only	Bracket Imputation Only	Model and Bracket Imputation	Valid Blank Converted
		Row percent distribution				
2010	14,296	46.3	25.2	19.2	6.0	3.3
2011	13,925	46.6	24.9	19.1	5.7	3.6
2012	13,761	47.9	24.8	18.8	5.0	3.5
2013	12,335	50.2	24.5	16.3	5.3	3.7
2014	13,305	50.0	23.2	18.1	5.3	3.4
2015	11,841	50.1	23.3	17.5	5.0	4.1

APPENDIX

Response and nonresponse rates

Appendix Table A shows the mapping of the CE final disposition codes to the American Association for Public Opinion Research (AAPOR) final disposition codes for in-person household survey:

Appendix Table A. Mapping of CE final disposition codes with AAPOR codes***

AAPOR(2015, p.67) Table 2: Final disposition codes for in-person, household survey		CEQ Final Disposition Codes (OUTCOME)	CED Final Disposition Codes (PICKCODE)
1. Interview	**1.0**		
Complete **(I)**	1.1	201 Completed interview	201 Completed interview
Partial **(P)**	1.2	203 Sufficient partial (through Section 20, no further follow-up)	*217 Interview- Temporarily Absent
2. Eligible, Non-Interview	**2.0**		
Refusal and break-offs (R)	2.10		
Refusals	2.11	321 Refused, hostile(A) 322 Refused, time(A) 323 Refused, language (A) 324 Refused, other (A)	321 Refused, hostile(A) 322 Refused, time(A) 323 Refused, language (A) 324 Refused, other - specify (A)
Household-level refusal	2.111	na	na
Known respondent refusal	2.112	na	na
Break-off	2.12	215 Insufficient partial (A)	
Non-contact (NC)	2.20		
Unable to enter building/reach housing unit	2.23	219 Other (A)	219 Other (A)
No one at residence	2.24	216 No one home	216 No one home
Respondent away/unavailable	2.25	217 Temporarily absent	
Other (O)	2.30	324 Refused, other	324 Refused, other
Dead	2.31	219 Other (A)	219 Other (A)
Physically or mentally unable/incompetent	2.32	219 Other (A)	219 Other (A)
Language (did not refuse)	2.33	219 Other (A)	219 Other (A)
Household-level language problem	2.331	219 Other (A)	219 Other (A)
Respondent language problem	2.332	219 Other (A)	219 Other (A)
No interviewer available for needed language	2.333	219 Other (A)	219 Other (A)
Miscellaneous	2.36	219 Other (A)	219 Other (A) 325 Diary placed too late (A)

Appendix Table A. Mapping of CE final disposition codes with AAPOR codes***

AAPOR(2015, p.67) Table 2: Final disposition codes for in-person, household survey		CEQ Final Disposition Codes (OUTCOME)	CED Final Disposition Codes (PICKCODE)
			326 Blank diary, majority of items recalled w/o receipts (A)
3. Unknown eligibility, non-interview **	**3.0**		
Unknown if housing unit occupied **(UH)**	3.10	na	na
Not attempted or worked	3.11		
Unable to reach/unsafe area	3.17		
Unable to locate address	3.18	258 Unlocated sample address (C): Treated as ineligible for CE	258 Unlocated sample address (C): Treated as ineligible for CE
Housing unit/Unknown if eligible respondent **(UO)**	3.20	na	na
No screener completed	3.21		
Other	3.90		
4. Not Eligible	**4.0**		
Out of sample	4.10		
Not a housing unit	4.50	228 Unfit, to be demolished (B) 229 Under construction, not ready (B) 231 Unoccupied tent/trailer site (B) 232 Permit granted, construction not started (B) 240 Demolished (C) 241 House/trailer moved (C) 243 Converted to permanent nonresidential (C)	228 Unfit, to be demolished (B) 229 Under construction, not ready (B) 231 Unoccupied tent/trailer site (B) 232 Permit granted, construction not started (B) 240 Demolished (C) 241 House/trailer moved (C) 243 Converted to permanent nonresidential (C)
Business, government office, other organization	4.51	243 Converted to permanent nonresidential (C)	243 Converted to permanent nonresidential (C)
Institution	4.52	Na	Na
Group quarters	4.53	252 Located on military base or post (C)	252 Located on military base or post (C)
Vacant housing unit	4.60	226 Vacant for rent (B) 331 Vacant for sale (B) 332 Vacant other (B) 341 CU moved (C) 342 CU merged with another CE CU within the same address (C)	226 Vacant for rent (B) 331 Vacant for sale (B) 332 Vacant other (B) 341 CU moved (C) 342 CU merged with another CE CU within the same address (C)
Regular, Vacant residences	4.61		
Seasonal/Vacation/Temporary residence	4.62	332 Vacant other (B) 225 Occupied by persons with URE (B)	332 Vacant other (B) 225 Occupied by persons with URE (B)
Other	4.63	233 Other (B)	233 Other (B)

Appendix Table A. Mapping of CE final disposition codes with AAPOR codes***			
AAPOR(2015, p.67) Table 2: Final disposition codes for in-person, household survey		**CEQ Final Disposition Codes (OUTCOME)**	**CED Final Disposition Codes (PICKCODE)**
		244 Merged units within same structure (C) 245 Condemned (C) 247 Unused serial number or listing sheet (C) 248 Other (C) 259 Unit does not exist or is out of scope 290 Spawned in error (C)	244 Merged units within same structure (C) 245 Condemned (C) 247 Unused serial number or listing sheet (C) 248 Other (C) 259 Unit does not exist or is out of scope
No eligible respondent	4.70	224 All persons under 16 (B)	224 All persons under 16 (B)
Quota filled	4.80	Na	Na

NOTES:
Census Bureau non-interview categories: (A)=Type A (B)=Type B (C)=Type C
* CED: Type A code "217 – temporarily absent" is treated as "completed interview" by CE-SMD. The Diary survey is designed to collect data for respondents when they are at home, and the Interview survey is designed to collect data for respondents when they are both at home and away on trips. When everyone is away on a trip in a Diary household for the entire week, they are counted as completed interviews with $0 of expenditures at home. Instead, expenditures for those away on trips comes from the Interview survey. Since Diary and Interview data are merged or "integrated" during estimation, this practice is designed to capture the right amount of expenditures.
** CE does not have an "Unknown eligibility" classification because Census trains interviewers to treat any case of unknown eligibility as Type A. This is consistent with AAPOR recommendation: if the definitive situation for a case cannot be determined, one should take the conservative approach of assuming the case is eligible or possibly eligible rather than not eligible.
***Reference: The American Association for Public Opinion Research (2015). *Standard Definitions: Final dispositions of case codes and outcome rates for surveys*. 8th edition.

In the following definitions for eligible sample, response rate, refusal rate, noncontact rate, and other non-response rate, the formula contain the alphabets I, P, R, NC, O, which refer to groupings of final disposition codes that are defined in Appendix Table A above.

Eligible Sample (denominator for response, refusal, noncontact, and other nonresponse rates)
= I + P + R + NC + O
The total number of eligible units - those who completed interviews (I, P), plus non-response due to refusals, non-contact, or other reasons (R, NC, O). This excludes any address that was sampled and ineligible (for example, an abolished household at a sampled address or a commercial business at a sampled address).

Response Rate (AAPOR definition RR2)
= (I + P) / (I + P + R + NC + O)
Defined as total number of good and partial interviews (interviews that provide data for use in the production tables), divided by the eligible sample. For the CE, unknown eligible housing units are coded as "Eligible non-interview" (i.e. Type A).

Refusal Rate (AAPOR definition REF3)
= R / (I + P + R + NC + O)
Defined as total number of eligible non-responses that were refused or started, but not completed, divided by the eligible sample. Refused interviews includes refusals due to time, language problems, and other types of refusals.

Noncontact Rate (1 - AAPOR definition CON3)
= NC / (I + P + R + NC + O)

Defined as total number of eligible non-responses due to inability to make contact with an eligible sample unit member.

Other Nonresponse Rate
= O / (I + P + R + NC + O)
Defined as total number of eligible non-responses due to reasons other than refusal and noncontact with an eligible sample unit member.

The sum of Response Rate, Refusal Rate, Noncontact Rate, and Other Nonresponse Rate comprise 100 percent of the universe of eligible sample units. In addition to these four rates, we also report on the Nonresponse Reclassification rate, which is a subset of Other Nonresponse cases.

Nonresponse reclassification rate
Defined as the total number of interviews that were changed from completed to a Type A non-interview based on a review of total expenditures (CE's Minimal Expenditure Edit routine) and other information about the CU, divided by the eligible sample.

For the CEQ
OUTCOME = 219 Other Type A Noninterview, & TYPEASP = "Minexpn"

For the CED

INTRVIEW =

5 Diaries with zero items reported in both weeks of the survey OR Diaries with zero items reported and the diary from the other diary week is a Type A, B, or C non-interview

6 Diaries with zero items reported and the diary from the other diary week has > 10 items reported in FDB with the total cost of these items being <= $50 OR Diaries with zero items reported and the diary from the other diary week has <= 10 items reported in FDB with the total cost of these items being <= $50 and the CU does not live in a rural area or a college dormitory and no members of the CU were away during the reference period

7 Diaries where there is one person in the CU and the total amount spent on food (at home and away from home) is <= $5 in the current week and <= $15 in the other diary week, and the number of items reported for non-food items in the current week is < 4 or the total cost of items reported for non-food items in the current week is < $30

8 Diaries where there are 2 or 3 members in the CU the total amount spent on food (at home and away from home) is <= $10 in the current week and <= $20 in the other diary week and the number of items reported of non-food items in the current week is < 4 or the total cost of non-food items reported in n the current week is < $30

9 Diaries where there are four or more CU members and CU the total amount spent on food (at home and away from home) is <= $20 in the current week and <= $30 in the other diary week and the number of items reported of non-food items in the current week is < 4 or the total cost of non-food items reported in n the current week is < $30

Summary of changes to data collection in 2015

In this section, we provide a brief overview of the changes in data collection in 2015 that impact the universe of eligible sample units included in the production of CE's official published tables, and response rate computations.

1. CE Sample Redesign

The CE sample is updated after every Decennial Census to ensure it reflects the population. The 2010 Decennial Census geographic boundaries were implemented for the CE in 2015, and are henceforth referenced as the *2010 Sample Redesign*. The first month of expenditures in 2015 that are eligible under the 2010 Sample Redesign is *January 2015*.

- CEQ: The CEQ has a three-month *retrospective* reference period *prior* to the month of data collection (sample month). Thus, *February 2015* is the first sample month for CEQ cases under the 2010 Sample Redesign for producing the official published tables for 2015 (since the February 2015 sample month has a reference period of Nov 2014, Dec 2014, and *Jan 2015*).
- CED: Unlike the CEQ, the CED has a *prospective* 1-week reference period after the diary is placed. Thus, *January 2015* is the first sample month for CED cases under the Sample Redesign for producing the official published tables for 2015.

2. Bounding Interview dropped in the CEQ

The CEQ bounding interview in Wave 1 of the five-wave survey panel was dropped starting with the 2010 Sample Redesign. The bounding interview had a 1-month recall and its data had not been previously used to produce estimates for the CE official published tables. Thus with the dropping of the bounding interview, cases in the 2015 CEQ survey panel under the 2010 Sample Redesign, and moving forward, will comprise of four waves of interview. However, due to the rotating panel design of the CEQ, there were still Wave 5 cases from CEQ survey panels that started in 2014 but did not complete until 2015.

In summary, the CE data used to compute final disposition rates to match the data used in the production of CE's official published tables are as follows:

CEQ	Prior to 2015	2015
Calendar months of data used	Jan through Dec	Feb through Dec
Waves in survey panel	2 through 5 (Wave 1 was bounding interview)	All waves (No bounding interview)

CED	Prior to 2015	2015
Calendar months of data used Diary week	Jan through Dec 1 and 2	Jan through Dec 1 and 2

Records use comparison groups

Respondent use of records for reporting expenditures is an indicator of accurate reporting of the amount of the expense and the details about the expense. At the end of the CEQ, there is a "Post Interview for Field

Representatives" section. One of the questions in that section asks the interviewer, "How often did the respondent consult records?" in the interview, and provides four response options:

1. always or almost always (90% of the time or more)
2. most of the time (50 to 89%)
3. occasionally (10 to 49% of the time)
4. never or almost never (less that 10% of the time).

Two comparison groups for records usage was created for analysis, "*Records*" vs "*None*":

- Records group consisted of interviews in which the interviewer reported records were used occasionally, most of the time, or always;
- None group comprised of interviews in which the interviewer reported records were used never or almost never.

Survey mode comparison groups

Although the CEQ is designed to be a personal visit interview, phone interviews may be conducted where it would be costly in time and resources to travel to a CU. However, a personal visit is the preferred method of data collection because of the opportunity for the interviewer to build a rapport with the respondent, and thereby hopefully increase the likelihood of encouraging respondents to use records. Since use of records is highly desirable for meeting the objective of accurate reporting, it is important to know if records use differs by survey mode. In the Post Interview for Field Representatives section, the interviewer is also asked about the mode used to collect data for the interview, and 6 response options are provided:

i. Personal visit for all sections,
ii. Personal visit for all sections, but telephone follow-up for some questions
iii. Personal visit for more than half of the sections, the rest by telephone
iv. Equally split between personal visit and telephone
v. Telephone for more than half of the sections, the rest by personal visit
vi. Telephone for all sections

Three comparison groups for survey mode was created for analysis, *"Personal Visit", "Phone",* and *"Mixed*":

- *Personal Visit* group consisted of interviews reported to be conducted by modes i. and ii
- *Phone* group consisted of interviews by mode vi
- *Mixed* group consisted of interviews by modes iii., iv, and v.

Processed expenditures, Reported item counts

Processed expenditure records

For the CEQ. The processed expenditure records file is sourced from the CE Production database Post Phase 3 MTAB data table for the CEQ.

For the CED. The data source is the CE Production database Post Phase 3 EXPN data tables (EFDB, EMLS, ECLO, EOTH), excluding records with invalid cost values

*Reported item counts***.**

Ideally, the reported expenditure records files would comprise the expenditure variables that correspond directly to the survey questions about the expenditures. However, because there are hundreds of expenditure variables spread across more than 40 data tables, for convenience, modifications were made to the MTAB data to create the reported expenditures file. The following paragraphs describe the creation of the reported expenditures file for the CEQ and the CED.

For the CEQ. We made modifications to the MTAB file to attempt to get as accurate a count of the number of uniquely reported items for each CU as possible while still taking advantage of the convenience of the MTAB data file. We subset the MTAB data file to records that were unique by the combination of three variables on the MTAB file: CU identifier (FAMID or NEWID), SEQNO, and EXPNAME. Then, the COST_ flag associated with each record was used to determine the type of data edit for each record. We acknowledge these modifications are not comprehensive enough to capture all post-data collection edits (for example edits made to a source variable may not carry forward to the mapped variable and edits made to non-cost fields are not captured), but these modifications make some strides towards that goal.

*For the CED***.** We made modifications to the Post Phase 3 EXPN files by extracting records that were unique by the combination of two variables on the EXPN files: CU identifier (FAMID or NEWID) –SEQNO. Then, the flag variable COST_ was used to determine the type of edit. Again, this modification does not capture the universe of all edits made in processing, but it does improve the accuracy of our computed edit rates relative to what has been reported previously.

Count of reported (expenditure) items per CU. The number of reported expenditure items by the CU was computed by counting the number of records in the *reported expenditures file* for a CU.

Expenditure edit type

CE Interview Survey (CEQ)

Interview expenditure edits are calculated using the interview monthly tabulation file (MTAB). The flag variable COST_ is used to identify if an expenditure was edited and what type of edit was done (imputation, allocation, combination, other). In addition, the "allocation number" is used to determine whether the resulting estimate has been allocated. The different types of edits (or non-edits) was identified by the following flag values for the CEQ:

CEQ MTAB Flag value	**Flag Description**	**Edit group**	**Edit Subgroup**
0	All of the source fields were flagged either as 0 (No Census adjustment) or -300 output from screens selected for microfilm review/no change or -400 output from screens; but not selected for microfilm review (no change)	Unedited	NA
1	One of the source fields was flagged by Census (source flag >0)	Unedited	NA
2	Manually updated (expenditure flag = -100) Changed in superfix (not a valid data adjustment source record field [-500]) Changed in superfix (is a valid data adjustment source record field [-600]) (Note: All of the following flags (3-9 & Q-S) indicate the source field was data adjusted by BLS. The two digit numbers in the parenthesis are the trailing digits of the source field flag, and indicate the method(s) of adjustment named after the parenthesis.)	Edited	Other
3	(-01 through -10) IMPUTATION	Edited	Imputed
4	(-12 through -19) ALLOCATION	Edited	Allocated
5	(-20 through -27) IMPUTATION and ALLOCATION	Edited	Combination
6	(-30 through -32) COMPUTATION only	Unedited	NA
7	(-35 through -43) COMPUTATION and IMPUTATION	Edited	Imputed
8	(-45 through -52) COMPUTATION and ALLOCATION	Edited	Allocated
9	(-53 through -68) COMPUTATION, IMPUTATION and ALLOCATION	Edited	Combination
Q	(-70 through -74,-75,-76) MANUAL IMPUTATION	Edited	Imputed
R	(-78 through -85,-86,-87,-88) MANUAL ALLOCATION	Edited	Allocated
S	(-90) SECTION 18 SPECIAL PROCESSING	Edited	Other

CE Diary Survey (CED)

The diary expenditure edit rate is calculated using the expenditure files from diary. The flag variable COST_ is used to identify if an expenditure was edited. In addition, the "allocation number" is used to determine whether the resulting estimate had been allocated. An expenditure record will be considered unedited if it has one of the following flags:

CED EES Flag Value	Description	Explanation
'0'	Default - no change to data	No adjustments were made during processing.
'-3'	Reviewed, no update; default adjustment status	The value was reviewed during processing, but no adjustments were made.
'11'	Sales Tax, Preliminary edits, or Minimal expenditure reclassification edit	Sales tax is a calculation applied to the data and will be treated as unedited for these rates.
'15'	Phase 1 Confirmed. Operator/Error Resolution Overrides (confirms value)	This flag is carried from the CAPI instrument and is present when a Field Representative suppresses a prompt to check the value (confirming the reported value). No changes are made to the data.
'16'	Phase 1 Changed. Error Resolution Changes value	This flag is carried from the CAPI instrument and is present when a Field Representative updates a value after prompted to check the value. Though the data is changed, it is assumed that it is edited based on the respondent's input and not considered as edited during processing.

All other flags indicate some type of adjustment during processing and are considered edited. An allocation rate is also produced using the allocation number of a given item (ALCNO). Any allocation number not equal to '000' is an allocated value. It is important to note that the values that are allocated are included in the editing rate; however, these values may also have been edited in some other way during the processing. It is not possible to delineate other edits from the current data available.

Note: for both CEQ and CED, the number of targets selected for an allocation will affect the adjustment rates - the total number of items that are allocated will add to both the numerator and the denominator for analysis using *processed* expenditure records (but not reported expenditure records).

Income edit type

The CE implemented multiple imputations of income data, starting with the publication of 2004 data. Prior to that, only income data collected from complete income reporters were published. However, even complete income reporters may not have provided information on all sources of income for which they reported receipt. With the collection of bracketed income data starting in 2001, this problem was reduced but not eliminated. One limitation was that bracketed data only provided a range in which income falls, rather than a precise value for that income. In contrast, imputation allows income values to be estimated when they are not reported. In multiple imputations, several estimates are made for the same CU, and the average of these estimates is published.

Income data from the Diary Survey are processed in the same way as in the Interview Survey.

Imputation rates for income are calculated based on the processed CE data (EES Post Phase 3 data that are used to produce the published tables) for each collection period. Following the model of the production tables, each wave of data will be treated independently for the CE quarterly interview survey (CEQ) and each weekly diary are treated independently for the Diary survey (CED). Imputation rates are calculated for final income before taxes. The income is counted as imputed if any of its summed components were imputed during processing. This will be identified using the imputation indicator flag. Any value of the flag not equal to '100' is considered imputed.

Imputation Flag Value	**Description**
100	No imputation. This would be the case only if NONE of the variables that are summed to get the summary variables is imputed.
2nn	Imputation due to invalid blanks only. This would be the case if there are no bracketed responses, and at least one value is imputed because of invalid blanks.
3nn	Imputation due to brackets only. This would be the case if there are no invalid blanks, and there is at least 1 bracketed response
4nn	Imputation due to invalid blanks AND bracketing
5nn	Imputation due to conversion of valid blanks to invalid blanks. (Occurs only when initial values for all sources of income for the consumer unit and each member are valid blanks.)

NOTE: an all valid blank conversion rate is calculated indicating the percent of instances that were converted from all valid non-responses (i.e., the respondent replied that the CU did not receive income from any source) to invalid non-responses that were subsequently imputed during processing. This will be based on the indicator flag with a value of '500' or above.