



# COVID-19 Response and Data Quality Impacts in the CE Interview Survey

October 2021

Yezzi Angi Lee & David Biagas  
*Consumer Expenditure Surveys Program Report Series*

## I. Introduction

The COVID-19 pandemic placed new challenges on the Bureau of Labor Statistics' data collection efforts. For example, with personal contact restricted in accordance with health policy or law, surveys traditionally collected in-person have had to rely on other modes of collection (e.g., telephone). This report describes how the pandemic-induced changes in collection methods have affected data quality in the Interview component of the Consumer Expenditure Surveys (CE).

The CE is comprised of two household surveys, the Interview Survey and the Diary Survey, that collect expenditures, income, and demographic characteristics of consumers in the United States and are collected by the U.S. Census Bureau on behalf of the Bureau of Labor Statistics. The Interview Survey, which is the focus of this report, is a rotating, four-quarter panel (wave 1 through wave 4) designed to collect large and recurring expenditures that consumers can be expected to recall for a period of three months or longer. The Diary Survey, the subject of a different report on quality during the COVID-19 pandemic<sup>1</sup>, is a two-week survey in which respondents record daily expenditures and is particularly important for collecting data on frequently purchased or “small-ticket” items for which expenditures are not likely to be recalled for long periods (e.g., detailed food expenditures, such as white bread or lettuce, and personal care products, such as hair care products and shaving needs).

---

<sup>1</sup> For a report on the effects of the pandemic on Diary Survey data, see McBride and Graf (2021).

The Interview Survey is designed to be an in-person interview. Although telephone interviews are allowed to accommodate the respondents' situations and to avoid refusals, they are generally limited to ensure better data quality. Between 2018 and 2019, the percent of wave 1 responding consumer units<sup>2</sup> completed via personal interview in a quarter ranged between 71.9 and 77.6 percent and the corresponding percent of wave 2 through 3 responding consumer units ranged between 57.9 and 60.8 as reported in the CE Data Quality Profile (DQP) (Knappenberger et al., 2021).

However, the COVID-19 pandemic placed new challenges on CE's data collection efforts, causing an abrupt change in procedures. For example, in response to the pandemic, the Census Bureau stopped all personal visits in mid-March 2020 for the health and safety of both interviewers and respondents and shifted to telephone interviewing for the CE. In addition, the Census Bureau temporarily suspended mailing out advance letters notifying households of their selection for the survey. As a result, nearly all interviews were collected by telephone<sup>3</sup> regardless of wave in the middle of March 2020.

The DQP provides overall estimates of data quality metrics (e.g. response rates, information book usage rates, respondent burden etc.) over time, up to and including the onset of the COVID-19 pandemic. In this research, we focused on the same data quality metrics to investigate the effect of the COVID-19 on Interview Survey's data quality, but we attempt to isolate data quality changes due to the change in protocol and not the pandemic. As part of this, we look at respondent demographics and data quality metrics by changes in mode of the interview. Note that, like the DQP, this report is a descriptive study, and we provided these data for the reader to interpret. Therefore, we attempted to move away from performing complicated statistical analysis. In a forthcoming report, we examine the effect of the COVID-19 pandemic using a discontinuous growth curve model to control for consumer unit characteristics and the changes in mode collection.

---

<sup>2</sup> Although a household refers to all people who live together in the same living quarters, "consumer unit" refers to the people living therein who are a family, or others who share in specific financial arrangements. For example, two roommates living in an apartment constitute one household. However, if they are financially independent, they each constitute separate consumer units within the household. Similarly, although families are related by blood, marriage, or legal arrangement, unmarried partners who live together and pool income to make joint expenditure decisions constitute one consumer unit within the household. For a complete definition, see the CE glossary at <https://www.bls.gov/cex/csxgloss.htm>.

<sup>3</sup> According to the DQP report (Knappenberger et al., 2021), 98 percent of respondents were interviewed by telephone. (See page 27)

## II. Data and Methods

We used data from July 2019 to June 2020, restricted to those who completed at least two interviews with mode information recorded<sup>4</sup> resulting in a sample size of 13,246. Table 1 shows the counts of completed interviews by month.

Table 1. Sample of consumer units by time period  
(N=13,246)

Type	Month	Count
	2019	
	July	1,182
	August	1,110
	September	1,084
	October	1,168
Pre COVID-19	November	1,123
	December	1,088
	2020	
	January	1,194
	February	1,089
	March	1,065
	April	1,068
During COVID-19	May	1,051
	June	1,024

### 1. Interview Schedule

The Interview Survey covers three months of expenditures, with each consumer unit interviewed up to four times over a period of consecutive quarters. For example, consumer units that were interviewed in January of a given year (wave 1), are interviewed again in April (wave 2), July (wave 3), and October (wave 4); whereas consumer units interviewed in February are interviewed again in May, August, and November. Each wave collects the past three months of expenditures, excluding the interview month. (That is, those interviewed in January report expenditures in October, November, and December of the prior year.) This study evaluates data quality across all four waves of data collection starting in July of 2019 through June of 2020 (Figure 6 in Appendix).

<sup>4</sup> There were 20,645 completed cases from July 2019-June 2020. Of those who completed the interview, 7,326 were removed because they had only one interview wave during the time period. An additional 73 cases were removed because the mode of interview was missing.

## 2. Definitions

We define Pre-COVID-19 as the interval from July 2019 through March 2020, a period before the start of the COVID-19 pandemic. *During COVID-19* is a time period after the start of the COVID-19 pandemic and consists of data collected from April 2020 to June 2020. The change of collection methods was defined by using the respondent's mode of completion from the last completed wave and compared with the mode of completion for the current wave. So, if a respondent completes wave 1 in-person and wave 2 over the telephone, then that respondent would be in the *in person - telephone* group in wave 2. If they refuse to participate in wave 3, but return to the survey in wave 4 via the telephone, then they would be classified in the *telephone-telephone* group for wave 4 (and excluded from the sample in wave 3). The distribution of change of collection methods are shown in Table 2. Lastly, Table 3 shows counts of change of collection methods by prior to and during the COVID-19 pandemic. For the purpose of this study, we will remove 45 cases of *in-person to in-person* and 9 cases of *telephone to in-person* identified during the COVID-19 pandemic<sup>5</sup>.

Table 2. Count of change of collection methods (N=13,246)

Type	Count
In-person to In-person	5,027
Telephone to Telephone	4,462
In-person to Telephone	3,287
Telephone to In-person	470

Table 3. Count of change of collection methods by time period (N=13,246)

COVID-19	Collection Changes	Count
Pre COVID-19	In-person to In-person	4,982
	Telephone to Telephone	3,050
	In-person to Telephone	1,610
	Telephone to In-person	461
During COVID-19	In-person to In-person	45
	Telephone to Telephone	1,412
	In-person to Telephone	1,677
	Telephone to In-person	9

<sup>5</sup> Although the CE survey changed to all telephone interview during the COVID-19 pandemic, there were 54 cases in total that were recorded as in-person interview during COVID-19. One possible explanation for these cases is interviewers misrecorded the mode of collection as in-person when it was a telephone interview.

### 3. Purpose Statement

The purpose of this study is to compare the data quality metrics of the Interview Survey before and after the onset of the COVID-19 pandemic. Earlier research has found differences in reporting behavior and data quality based on mode of data collection (Biagas 2020). Since the Census Bureau switched the method of collection to all telephone interviewing in March 2020 in response to the COVID-19 pandemic, we examine whether or not the change in the collection methods affected the quality of data. However, due to the pandemic, it is inherently difficult to isolate changes in reports resulting from the economic shock from changes due to differences in mode of collection. For example, changes in spending behavior due to the pandemic would result in a change in total expenditures, but past research has demonstrated that reports of total expenditures are also associated with differences in mode. We attempt to isolate the changes resulting from the pandemic by treating the *telephone to telephone* group as a control group since it was not impacted by the change of collection methods due to the COVID-19 pandemic. We make a naïve assumption<sup>6</sup> that respondents who remained in the telephone interview shows only the effect of the COVID-19 pandemic on changes in data quality and not the effect of a mode change.

One of the limitations of this study is we do not have respondent's true preference of mode (i.e. telephone or in-person) during the COVID-19 period. Since during the pandemic the CE Interview Survey was limited to telephone collection, *in-person to in-person* and *telephone to in-person* groups are missing during COVID-19 period. Therefore, we hypothesize that the majority of respondents who remained would have been classified as *in-person to in-person* in the pre COVID-19 period would likely fall in the *in-person to telephone* group in the during COVID-19 period along with those that would have naturally been in the *in-person to telephone* group as shown in Figure 1. We expect the *telephone to telephone* group to remain unchanged over the time periods. Therefore, we focus on the data quality trends of the *in-person to telephone* group during the COVID-19 period from *in-person to in-person* and *in-person to telephone* groups in the pre COVID-19 period.

---

<sup>6</sup> In the pre-pandemic period, there was 4.56 percent of consumer units who changed their mode from telephone to in-person.

Figure 1. Hypothesis



#### 4. Demographic Characteristics

Table 4 displays the demographic and socio-economic characteristics by change of collection methods in the pre and during COVID-19 periods. A study done by Biagas (2020) found that telephone respondents tend to have higher salaries and level of education than in-person respondents. We find similar results, as the *telephone to telephone* respondents from pre and during COVID-19 periods reported higher income and own higher level of education. Next, the *in-person to telephone* respondents were more likely to be older, white, and homeowners. We hypothesize the majority of respondents from *in-person to in-person* group in the pre COVID-19 period would likely shift to the *in-person to telephone* group in the during COVID-19 period due to protocol change. Table 4 supports this hypothesis, as the trend for most demographic and socio-economic characteristics in the *in-person to telephone* group during the COVID period are either similar to or falls between the averages of the *in-person to in-person* group and the *in-person to telephone* group for pre COVID-19 period. When we assume that the *in-person to in-person* group during the COVID-19 period shifted to the *in-person to telephone* after the onset of the COVID-19 pandemic, the increase in age of respondents for the *in-person to telephone* group during COVID-19 period seems reasonable. The same reasoning can be applied for mean income, in that the decreases in the *in-person to telephone* group from pre to during COVID-19 period can be explained by the shift of the *in-person to in-person* group in the pre COVID-19 period to the *in-person to telephone* group during COVID-19.

Table 4. Sample characteristics by change of collection methods and time period

Demographic	Pre COVID-19				During COVID-19 <sup>7</sup>	
	In-person to In-person	In-person to Telephone	Telephone to In person	Telephone to Telephone	In-person to Telephone	Telephone to Telephone
N	4,982	1,610	461	3,050	1,677	1,412
Mean Age	56.1	52.7	52.8	52.4	56.7	53.3
(SE)	0.3	0.4	0.8	0.3	0.4	0.4
Mean Income	\$77,750.49	\$83,167.43	\$76,458.57	\$91,727.41	\$80,153.66	\$91,619.63
(SE)	\$3,172.47	\$2,205.94	\$3,677.77	\$1,772.53	\$2,048.45	\$2,423.63
Mean Family Size	2.4	2.5	2.5	2.3	2.4	2.4
(SE)	0.02	0.04	0.07	0.03	0.03	0.04
Respondent's Race*						
<i>White</i>	70.1%	69.2%	67.0%	70.3%	72.2%	70.9%
<i>Black</i>	8.9%	10.4%	11.1%	10.3%	8.4%	9.1%
<i>Asian Pacific</i>	4.9%	5.9%	5.0%	6.9%	4.4%	7.0%
<i>Hispanic</i>	14.4%	12.5%	15.2%	10.6%	13.5%	10.8%
<i>Other</i>	1.7%	1.9%	1.7%	1.9%	1.5%	2.2%
Respondent's Sex						
<i>Female</i>	50.7%	53.5%	52.5%	51.4%	50.7%	50.5%
<i>Male</i>	49.3%	46.5%	47.5%	48.6%	49.3%	49.5%
Respondent's Education*						
<i>Less than HS</i>	11.1%	8.1%	9.1%	7.0%	9.1%	6.7%
<i>High School</i>	24.5%	22.4%	22.6%	19.5%	25.5%	20.7%
<i>Some College</i>	20.9%	21.8%	21.9%	18.8%	19.4%	19.8%
<i>Undergraduate</i>	30.2%	33.7%	32.8%	37.7%	31.7%	35.4%
<i>Postgraduate</i>	13.3%	14.0%	13.7%	17.0%	14.3%	17.4%
Housing Tenure*						
<i>Owner</i>	66.7%	65.0%	64.6%	68.1%	71.8%	67.6%
<i>Other</i>	33.3%	35.0%	35.4%	31.9%	28.2%	32.4%

### III. Findings

To assess data quality changes, we compare the following factors before and after the onset of the COVID-19 pandemic: average total expenditures, survey length, record usage rates, information booklet<sup>8</sup> usage rates, number of entries, number of missing items, number of rounded items, edit flag rates, and perceived burden before and after COVID-19 data collection changes and based on mode of survey response. As shown in Table 5, the respondents who switched their mode from *in-person to telephone* during the COVID-19 period have the lowest median expenditures. Independent of time period, there was a high percentage of respondents who did not have any access to the information booklet for telephone interviews.

---

<sup>8</sup> Interview Survey respondents are supplied with an information booklet to assist them in answering survey questions. The information booklet provides examples that can clarify the kinds of expenditures that each section and item is intended to collect, and it also provides response options for demographic questions and bracketed response options.

Table 5. Data Quality metrics by change of collection method and time period

Demographic	Pre COVID-19				During COVID-19	
	In-person to In-person	In-person to Telephone	Telephone to In-person	Telephone to Telephone	In-person to Telephone	Telephone to Telephone
N	4,982	1,610	461	3,050	1,677	1,412
Mean Survey Length in min (SE)	65.3 0.5	57.8 0.7	59.2 1.5	57.7 0.5	59.9 0.7	60.0 0.8
Mean Expenditures (SE)	\$14,169.66 \$175.71	\$14,666.32 \$307.11	\$13,776.50 \$512.10	\$15,331.94 \$213.45	\$13,277.40 \$275.10	\$14,724.68 \$381.19
Mean Number of Entries (SE)	38.0 0.2	35.8 0.3	35.2 0.7	36.6 0.3	34.3 0.3	34.2 0.3
Mean # of Item-nonresponse (SE)	0.6 0.03	0.7 0.05	0.8 0.12	0.6 0.03	0.6 0.05	0.6 0.05
Mean # of Roundings (SE)	20.0 0.1	19.3 0.2	19.3 0.5	19.8 0.2	17.4 0.2	17.7 0.2
Mean Edit Rates (SE)	6.7 0.1	7.2 0.2	7.5 0.3	7.8 0.1	6.7 0.1	7.2 0.2
Perceived Burden						
<i>No Burden</i>	38.0%	26.4%	32.5%	24.2%	35.4%	28.1%
<i>Burden</i>	62.0%	73.6%	67.5%	75.8%	64.6%	71.9%
Infobook Usage						
<i>Did not use</i>	24.6%	2.6%	27.5%	2%	2.2%	0.8%
<i>Used</i>	57.2%	3.7%	42.1%	1.8%	3.5%	2.2%
<i>No Access</i>	18.2%	93.7%	30.4%	96.2%	94.4%	97.0%
Record Usage						
<i>Never Used</i>	45.5%	50.9%	52.3%	49.2%	47.9%	49.7%
<i>Used</i>	54.5%	49.1%	47.7%	50.8%	52.1%	50.3%
Income Impute						
<i>Not imputed</i>	62.5%	59.9%	57.4%	55.0%	62.5%	56%
<i>Imputed</i>	37.5%	40.1%	42.6%	45.0%	37.5%	44%

Next, we display the visual representations of data quality metrics (median total expenditures, median interview time, median number of entries, and median number of rounded entries) by collection changes across months as shown in Figures 2 through 5. Before we move on the result, the authors would like to acknowledge that a limitation of this study is that we do not know the respondent's true preference of the mode during the COVID-19 period, since the CE Interview Survey was limited to telephone collection. As we show in Figure 1, we hypothesize that the pre COVID-19 *in-person to in-person* group will likely fall into the *in-person to telephone* group during the COVID-19 period since in-person interviewing was restricted. In Figure 2, the median total expenditure for the *in-person to telephone* group decreases at faster rate than the *telephone to telephone* group. Since we assume that the respondents from *in-person to in-person* group before the COVID-19 pandemic are included in the *in-person to telephone* group during the COVID-19 period, when we consider the trend line for this group and the *in-person to telephone* group in the pre COVID-19 periods combined, then the trend seems more flattened. This suggests that overall, the data quality has not decreased. As shown in Figure 3, the median total interview time for respondents who remained *in-person to in-person* for pre COVID-19 period was almost always longer than all other groups. When the Census Bureau switched to all telephone interviewing after the start of the COVID-19 pandemic, the median interview time decreased for the *in-person to telephone* group. The trend for *telephone to telephone* group remains constant, which implies that the interview time was not much affected by COVID-19. Finally, the trend of the median number of entries (Figure 4) and median proportion of rounded values (Figure 5) remains relatively stable for both *in-person to telephone* group and *telephone to telephone* across the months.

Figure 2. Median Total Expenditures by collection changes

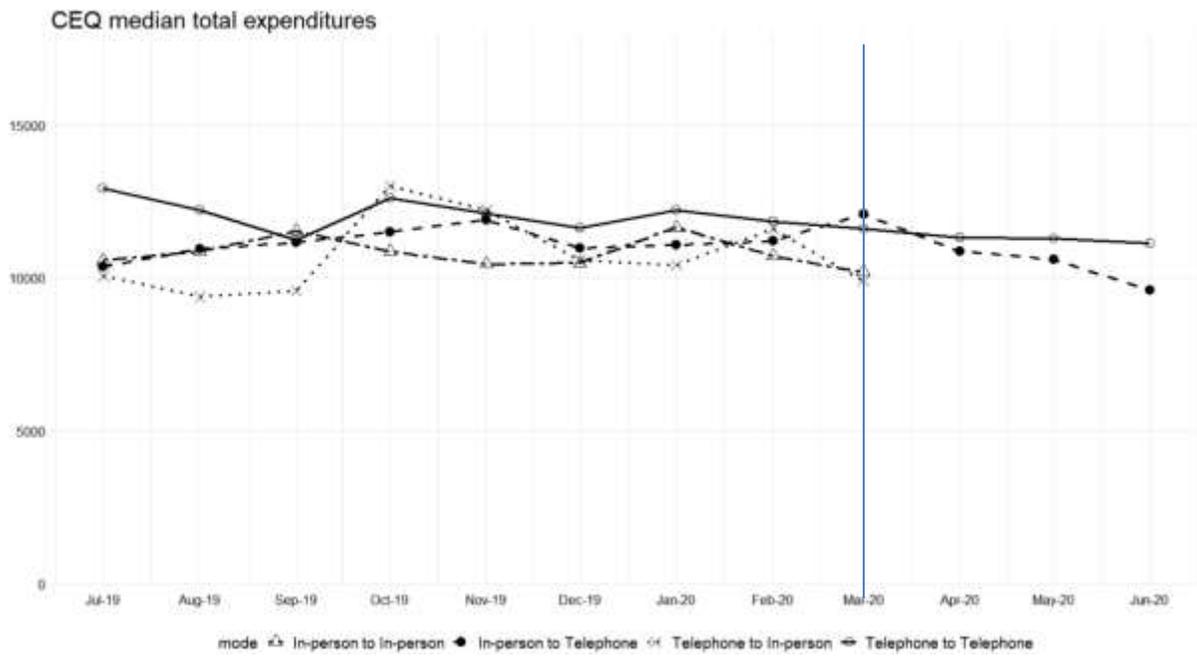


Figure 3. Median interview time by collection changes

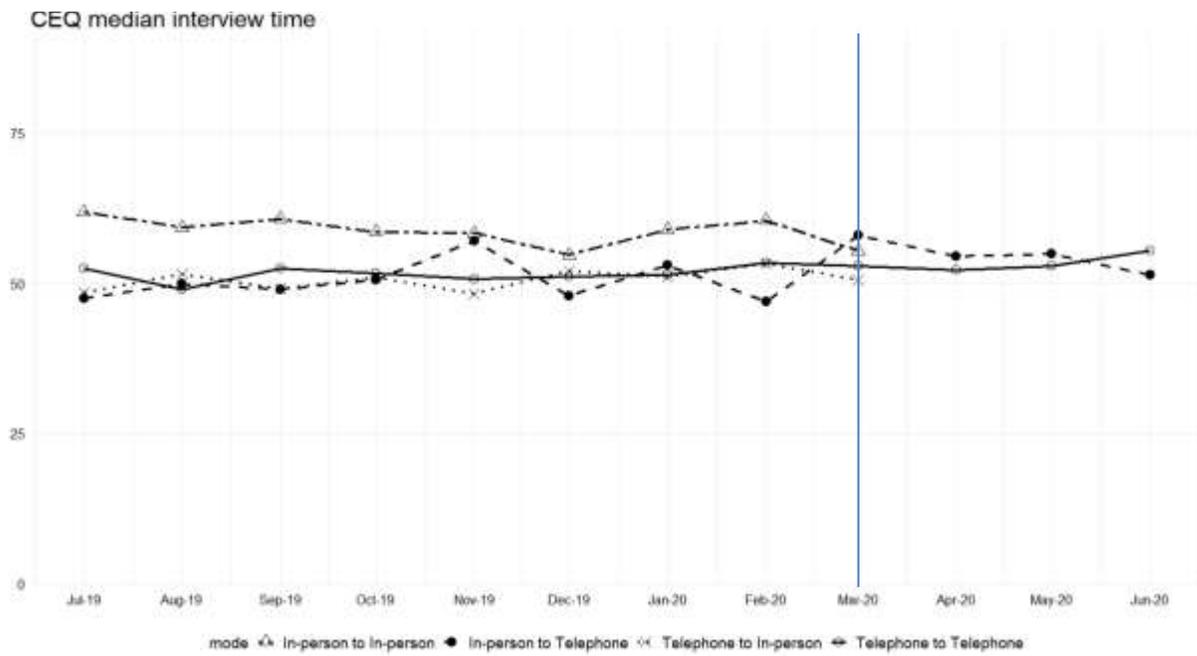


Figure 4. Median number of entries by collection changes

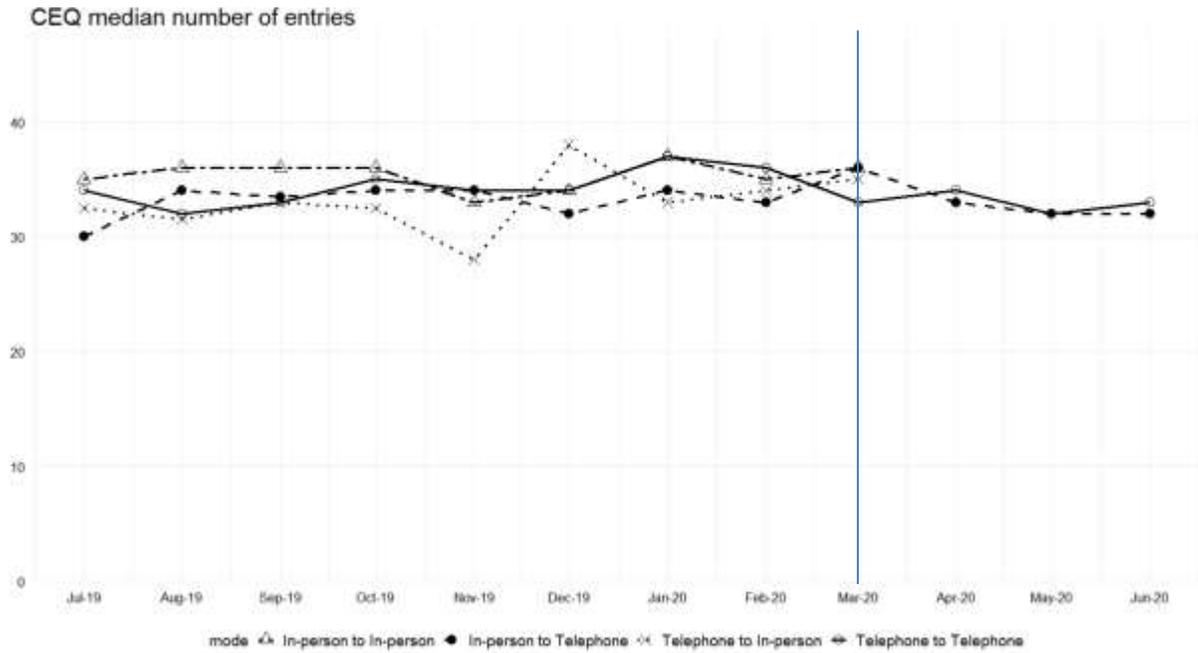
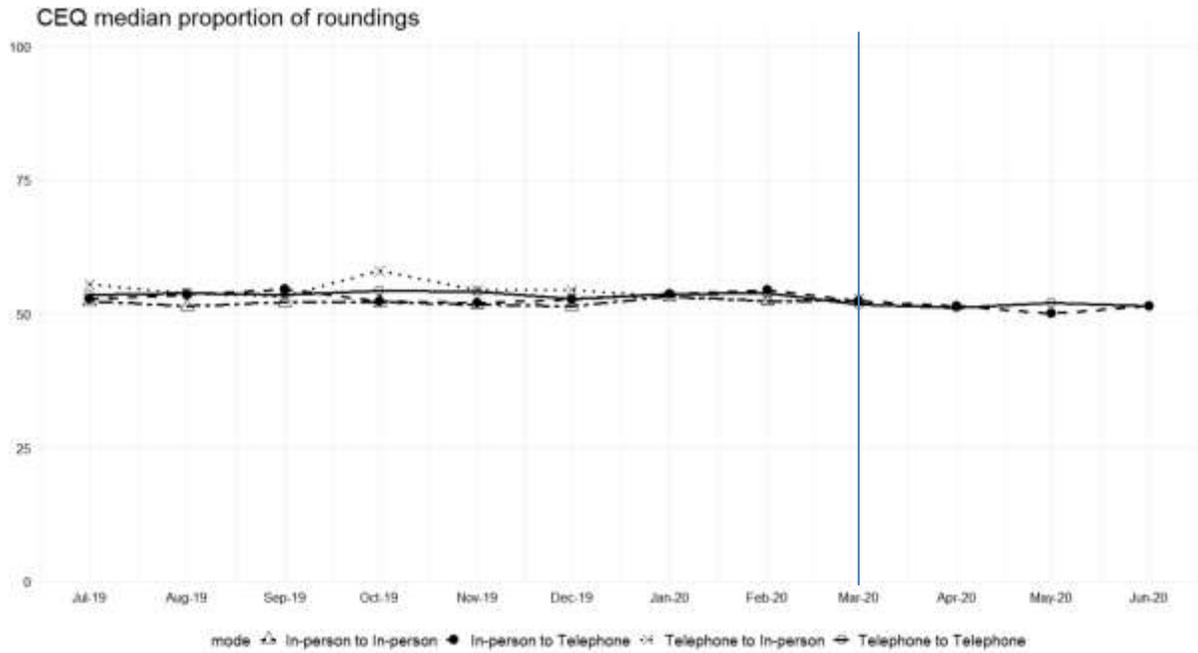


Figure 5. Median proportion of rounded values by collection changes



#### **IV. Conclusion**

In this research, we investigated the effect of the COVID-19 pandemic and the resulting mode changes on various data quality metrics in the CE Interview Survey. We discovered that respondents who switched from an *in-person to a telephone* interview after the start of the COVID-19 pandemic were more likely to be older, white, and homeowners and assumed these were those that would have otherwise remained in person prior to the pandemic. We also discovered the median total expenditures decreased for the *in-person to telephone* group across the months. However, when we considered that the *in-person to in-person* group in pre COVID-19 period shifted to the *in-person to telephone* after the onset of the COVID 19 pandemic, then the combined trend is flatter. The median survey length increased for the *in-person to telephone* group since the start of the COVID-19 pandemic while for the *telephone to telephone* group, the median was constant across the months. The median number of entries and the median proportion of rounded values displayed consistent trends across the months. There were substantial differences in the following data quality metrics between the collection change groups and the pre vs during COVID-19 periods: record usage rate and perceived burden, the number of entries, the number of missing entries, the number of rounded entries, survey length, total expenditures, information booklet usage rates and edit rates. (See table 5.)

Overall, we cannot draw a conclusion as to whether CE's data quality was affected by the change in mode due to COVID-19. However, it appears that the changes in data quality metrics were driven by the underlying characteristics of the respondents rather than the change in mode. In a forthcoming CE report, we will further look for differences in data quality resulting from the change in protocols due to the COVID-19 pandemic while controlling for consumer unit characteristics and the economic impact of the COVID-19 pandemic.

## V. Reference

- Knappenberger, Clayton, Yezzi Lee, Sharon Pham, and Grayson Armstrong. 2020. *The 2019 CE Data Quality Profile*. Consumer Expenditure Surveys Program Report Series. Washington DC: U.S. Bureau of Labor Statistics. September 9, 2021.
- McBride, Brett and Nikki Graf. *COVID-19 CE DIARY PRELIMINARY REPORT: Response and Data Quality Impacts*. Consumer Expenditure Surveys Program Report Series. Washington DC: U.S. Bureau of Labor Statistics. June 2021.
- Biagas, David. 2020. *Investigating the Effect of Increasing Telephone Completion Thresholds on the Reporting of Expenses on the Consumer Expenditure Survey*. Washington DC: U.S. Bureau of Labor Statistics

# VI. Appendix

Figure 6. Interview Wave by Month

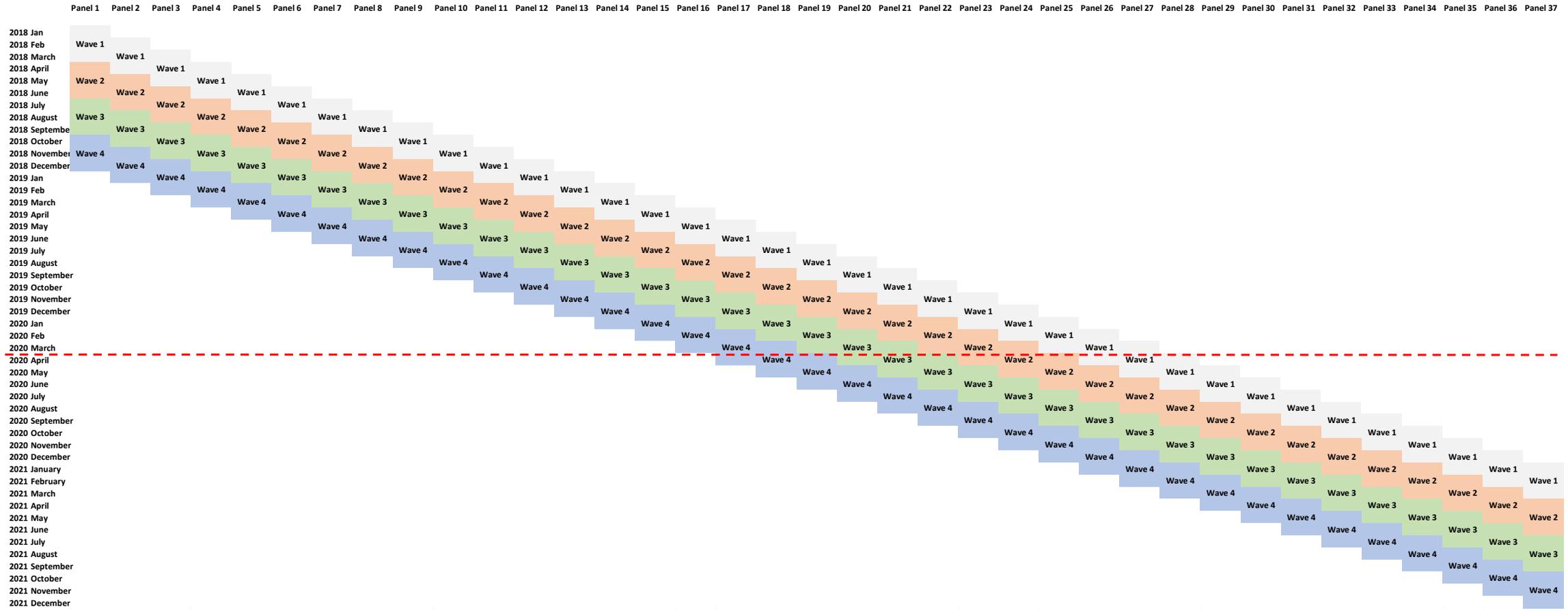


Table 6. Data Dictionary

Variable Name	Variable from CE Interview	Description
CUID Record Usage	NEWID RECORDS	Respondent indicator =1 if used record ( RECORDS = 1,2,3) =0 if did not used record (RECORDS = 4)
Number of rounded values		Sum of the counts of any reported value with 00,25,50 or 75 as the trailing digits for all expenditures
Number of entries		Sum of counts of reported expenditures
Number of item-nonresponse		Sum of the counts of reported expenditures with Don't knows or refusal
Family Size	FAM_SIZE	Number of Members in CU
Total Expenditures	ZTOTAL	Total expenditures
Information Booklet Usage	INFOBOOK	=1 if used infobook ( INFOBOOK = 1,2,3) =0 if did not used infobook/no access (INFOBOOK = 4,5)
Total Interview Time	TOT_TIME	Total Interview Time (min) TOT_TIME/60
Number of Edits	COST_	Sum of the counts of edits (COST_ =3,4,5,7,8,9,G,Q,R,S)
Income Imputation	FINCBTXI	=1 if imputed (FINCBTXI=100) =0 otherwise
Perceived Burden	RES01	=1 if reported any burden (RES01 = 2,3,4,5) =0 if reported no burden (RES01 = 1)
Wave	INTERI	
Month	QINTRVMO	Interview month
Age of reference person	AGE	Age of reference person (CU_CODE=1)
Highest education level of any member	EDUCA	Less than High School (EDUCA = 1,2,3) High School (EDUCA=4) Some College (EDUCA=5) Associate Degree or Bachelor's Degree (EDUCA=6,7) Graduate or Professional Degree (EDUCA=8)
Housing tenure and type of area	CUTENURE	=1 if owned (CUTENURE =1,2) =0 otherwise (CUTENURE = 3,4,5,6)
Income before taxes	FINCBTXM	Total amount of family income before taxes (CU_CODE=1)
Race of reference person	MEMBRACE HISPANIC	Hispanic (HISPANIC =1) White (MEMBRACE=1 & HISPANIC =0) Black (MEMBRACE=2 & HISPANIC =0) Asian and Pacific (MEMBRACE=4,5 & HISPANIC =0) Other (MEMBRACE=3,6,7,8,9 & HISPANIC =0)