Report on Nonresponse Bias during the COVID-19 Period for the Consumer Expenditures Diary Survey

November 1, 2022

Stephen Ash, Brian Nix, and Barry Steinberg

Consumer Expenditure Surveys Program Report Series

Contents

Overview

The emergence of COVID-19 in March 2020 caused extraordinary changes to many facets of life, and the operations of surveys such as the Consumer Expenditures Surveys (CE) are no exception. For example, the emergence of COVID-19 caused the Bureau of Labor Statistics to enact more restrictive interviewing protocols (Knappenberger et al., 2021), (Armstrong et al., 2022). These new interviewing protocols impacted response rates, and the CE's response rates decreased.

We define the COVID-19 period as the second and third quarters of 2020 (April through September of 2020) because most of the survey's data were collected by telephone interviewing rather than by in-person interviewing during that period. In mid-March 2020, the U.S. Census Bureau suspended all in-person interviews, and by April, close to 98 percent of all interviews were conducted over the phone regardless of wave (Knappenberger et al., 2021). This policy was gradually phased out in the spring of 2021.

Figure 1 presents overall response rates for the Diary Survey.

Figure 1: Diary Survey Response Rates

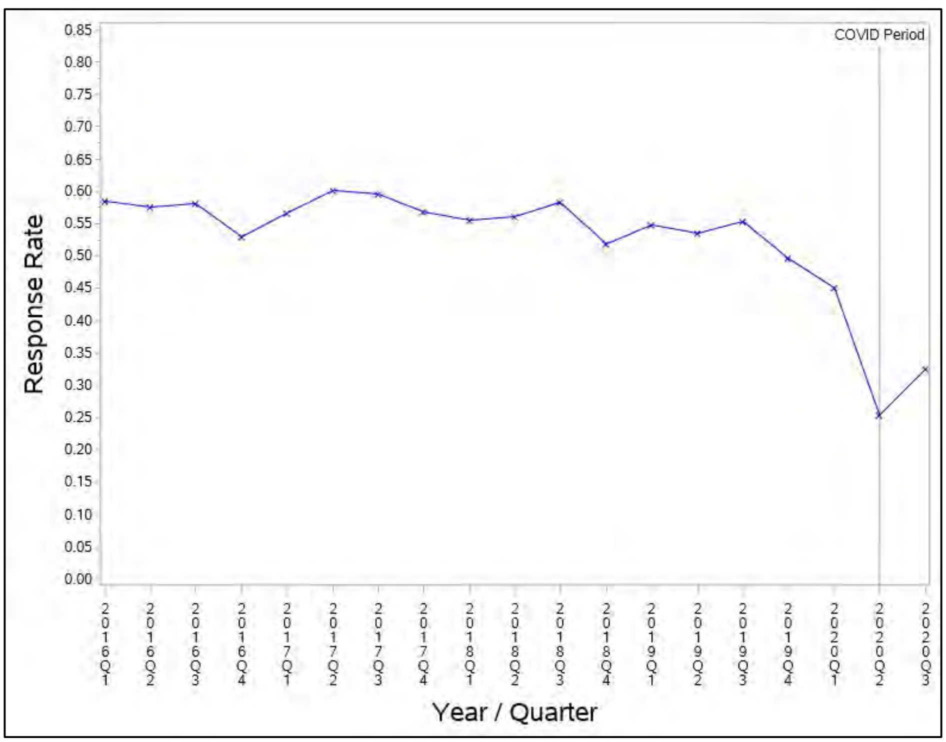


Figure 1 shows that the response rates fluctuated minimally between 2016Q1 and 2018Q3. There was a low of 53.0 percent in the 2016Q4, but rates rebounded thereafter and stayed between 51.8 and 55.4

percent until the 2019Q3 where the trend moved lower. The sharpest decrease of response rates coincides with the COVID-19 period.

Three studies have been released which examine the impact of COVID-19 on response rates and data quality in the CE: Lee and Biagas (2021) and Ash, Nix, and Steinberg (2022) examined the Interview Survey and McBride et al. (2021) examined the Diary Survey. This study is a companion to Ash, Nix, and Steinberg (2022) since it examines nonresponse bias with the Diary Survey and uses the same methods used by Ash, Nix, and Steinberg (2022) to examine the Interview Survey.

Prior comprehensive studies of nonresponse bias for the CE surveys include Chopova et al. (2008) and Steinberg et al. (2022) – they both found that nonresponse bias is relatively small, around one percent of the survey's published expenditure estimates. For more background on the CE surveys, see the *Handbook of Method*s (U.S. Bureau of Labor Statistics, 2022).

Summary of findings. We found no evidence of nonresponse bias in the expenditure estimates related to the COVID-19 Pandemic. Expenditure estimates were produced with three alternative weighting procedures that were tailored to the COVID-19 Pandemic, and all three expenditure estimates were in reasonable agreement with those produced with the normal weighting procedure used to produce public estimates.

The remainder of the report is organized as follows.

- The Technical Background defines nonresponse bias with respect to estimates and explains how the nonresponse adjustment in the weighting is used to reduce nonresponse bias. A summary is also provided of the nonresponse adjustment method currently used by CE weights.
- Analysis 1 uses graphs of response rates to review the variables used by the current nonresponse adjustment. This exploratory analysis considers whether the variables continue to have strong associations with response during the COVID-19 period.
- Analysis 2 provides measures of nonresponse bias by applying alternative nonresponse adjustments to expenditure estimates from the COVID-19 period.
- Analysis 3 compares estimates common to both CE and American Community Survey (ACS)[1] during from the COVID-19 period. Similar comparisons of 2018 and 2019 are provided as context.

[1] For more information about the American Community Survey, see the website at https://www.census.gov/programs-surveys/acs.

Technical Background

A brief description of CE's sample design and weighting procedures

In the CE Survey, expenditure data are collected from a representative sample of consumer units (CUs) across the nation.[2] The Census Bureau selects a representative sample of addresses for the survey[3]. Its field representatives then visit those addresses and attempt to collect expenditure data from the CUs that live there. Some addresses are *ineligible* for the survey because they do not meet eligibility criteria. Vacant housing units, businesses, and other non-residential addresses fall in this category. However, for addresses that have eligible CUs, the field representatives attempt to collect expenditure data from them. Some CUs participate in the survey, and some do not; we refer to CUs that participate in the survey as *respondents* and CUs that do not participate in the survey as *nonrespondents*.

Each CU in the sample has a *base weight* associated with it, which is the number of CUs in the population it represents. The sum of the base weights of all the CUs in the sample is the total number of CUs in the nation, which is currently around 132 million for collection year 2020. The base weights of the respondent CUs are increased to account for the nonrespondent CUs in a noninterview adjustment process. The nonresponse-adjusted weights are then adjusted to the number of "known" population totals from the Current Population Survey (CPS) in a calibration adjustment process. This report focuses on the noninterview adjustment process.

The noninterview adjustment process uses the cell adjustment method. The complete sample of CUs is partitioned into 192 cells according to: the region of the country in which they live, the number of people in their CUs, the number of contact attempts made by the survey's field representatives when trying to collect their data, and the average income in their ZIP code according to the Internal Revenue Service (IRS). The probability of a CU in the sample participating in the survey is estimated for each of the 192 cells by dividing the sum of the base weights from the respondent CUs by the sum of the base weights from all the CUs in the sample (respondents plus nonrespondents) within each cell. Then the weights of the respondent CUs are increased to account for the nonrespondent CUs by multiplying them by the inverse of their cell's estimated probability of participating in the survey. For technical background on the cell adjustment method, see also Oh and Scheuren (1983), Little (1986), and Brick (2013).

[2] A consumer unit (CU) is a group of people who live together and pool their incomes to make joint expenditure decisions. A CU can be thought of as similar to a "household." Ninety-nine percent of the time "consumer units" and "households" are the same thing, so the terms are often used interchangeably.

[3] The Bureau of Labor Statistics works with the Census Bureau to conduct the Consumer Expenditures Surveys. The Census Bureau provides the sampling frame, selects the sample, and conducts the interviews.

Nonresponse Bias and Weighting

The random response model (Bethlehem and Kersten 1985; Bethlehem, Cobben, and Schouten, 2011; p. 43) assumes that every CU in the sample has a unique probability of responding to the survey and does this by treating nonresponse as an additional stage of sampling (Särndal and Swensson 1987). That is, every CU has a probability of being selected for the sample, and every CU selected for the sample has a probability of responding to the survey. Those two random processes can be thought of as two phases of sampling. Then, in the same way that the sample weight is equal to the inverse of the probability of selection, the nonresponse weighting adjustment is equal to the inverse of the probability of response.

By treating the response to the survey as another stage of sampling, we define *nonresponse bias* as the impact of the imperfect nonresponse weighting adjustment on the estimates. If we knew the actual probability of response for every CU in the sample, we would produce unbiased estimates with respect to nonresponse. However, noninterview adjustments can only use the observed sample to estimate the probability of response, so the estimated probabilities of response always result in some level of nonresponse bias. The best we can do is to minimize the bias by having solid methods of estimating the probabilities of response, and then use this result to estimate the level of nonresponse bias.

In our study we measured the CE's nonresponse bias in the COVID-19 pandemic period as the difference between the expenditure estimates generated with the normal noninterview adjustment procedure, minus the expenditure estimates generated with noninterview adjustment procedures tailored to the COVID-19 pandemic period.

The current CE weights estimate the probability of response within cells using the cell adjustment method. The more alike the units within a cell are with respect to their probability of response, the closer the estimated probability of response will be to the actual probability of response, and therefore the less biased the survey's estimates will be with respect to nonresponse. However, we generally want to define cells with variables that are associated with both response and the variable of interest (Vartivarian and Little 2002). If cells are defined with variables that are associated with both, the survey's estimates will less biased with respect to both nonresponse and the variable of interest.

For CE, it makes sense to focus on forming nonresponse cells that are alike with respect to response because CE is used to produce a variety of estimates which makes it difficult to form cells with respect to every type of estimate. The variables used in the current noninterview adjustment for both the Diary Survey and the Interview Survey are identified by the last column in Table 1. The use of these variables is the result of the research of Dumbacher et al. (2012).

Table 1 also lists the variables considered in our research including **Error! Reference source not found.**the variable name, the source of the variable, and a short description. Our list is limited because the variables used with a noninterview adjustment need to be known for both respondents and noninterviews. This limitation often means that most of the variables used in noninterview adjustments are related to the sampling frame or are variables used in the administration of the survey. The variables from the Census Planning Database are described later in Analysis 2.

Table 1: Variables Considered for the Noninterview adjustment Research

Variable	**Source**	**Description**	**Current Method**
REGION	CE Sampling Frame	Census Region (4 categories)	X
DIVISION	CE Sampling Frame	Census Division (9 categories)	
WTIRSINC	Internal Revenue Service	IRS Income by ZIP code, (Top 10%, Middle 80%, bottom 10%)	X
WTNUMCNT	CE Administrative Data	Number of Contacts (2, 3-5,6-9, other)	X
MONTH	CE Interviewing Process	Month of Interview	
WTMEMQ	CE Interviewing Process	Number of persons in CU (1, 2, 3-4, 5+)	X
WTTENURE	CE Interviewing Process	Tenure (owner/renter)	
SR_NSR	CE Sampling Frame	Type of Primary Sample Unit (S=self-representing, N=non self-representing, not rural, R=Non Self-representing, rural)[4]	
Q1_65PLUS	Census Planning Database	Quartiles of percent of population within tract aged 65 years or older [Pop_65plus_ACS_09_13]	
Q1_Not_HS	Census Planning Database	Quartiles of percent of population within tract not a high school graduate [pct_Not_HS_Grad_ACS_09_13]	
Q1_BLW_POV	Census Planning Database	Quartiles of percent of population within tract below the poverty level [pct_Prs_Blw_Pov_Lev_ACS_09_13]	
Q1_MED_INC	Census Planning Database	Quartiles of tract-level median income [Med_HHD_Inc_ACS_09_13]	
Q1_BLACK	Census Planning Database	Quartiles of percent of population within tract Black Alone [pct_NH_Blk_alone_ACS_09_13]	
Q1_HISP	Census Planning Database	Quartiles of percent of population within tract Hispanic or Latino [pct_Hispanic_ACS_09_13]	

X – indicates the variable is included in the current noninterview adjustment of the Diary Survey and the Interview Survey.

[4] 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: 23 "S" PSUs, which are metropolitan CBSAs with a population over 2.5 million people (self-representing PSUs); 52 "N" PSUs, which are metropolitan and micropolitan CBSAs with a population under 2.5 million people (nonself-representing PSUs); and 16 "R" PSUs, which are non-CBSA areas ("rural" PSUs)

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. The CE uses all 91 of these PSUs in its 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.

Analysis 1: Exploratory Analysis of Response Rates

In this analysis, we review the effectiveness of the variables currently used in the noninterview adjustment process and output is shown in Figures 2 to 5 and Figures A1 to A10 of the Attachment. Effective variables stratify CUs into groups with different response rates, which can be seen in the graphs by lines that are far apart from each other. Effective variables stratify CUs into groups with similar response rates, which can be seen by lines that are close to each other and/or overlap each other. The more effective the variable, the farther apart the lines in the graphs. In this analysis, we are looking at the variable's effectiveness and whether their effectiveness changed in the COVID-19 period. The appendix includes similar graphs of response rates for the other variables considered in our analysis.

The nature of the analysis is exploratory and univariate. It is an exploratory analysis of the response rates in the sense that we are providing a graphical and intuitive understanding of the response rates including how the response rates have and have not changed over time with respect to their effectiveness in our noninterview adjustment. The analysis is also limited because it is univariate and not multivariate. We are only considering each variable one-at-a-time and not comparing their relationship to response altogether as in Analysis 2. Although there are no formal statistical tests associated in this section and the analysis is univariate, we think the graphs provide a useful background for an understanding of the response rates.

Figure 3 shows response rates by number of persons in the CU.

Figure 2: Response Rates by Number of Persons in the Consumer Unit

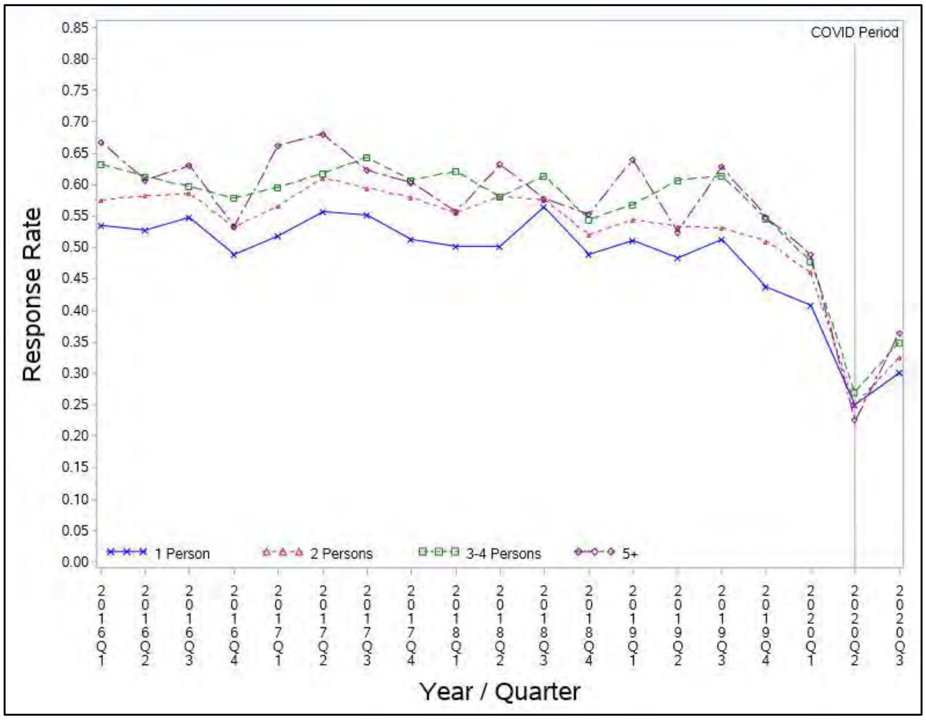
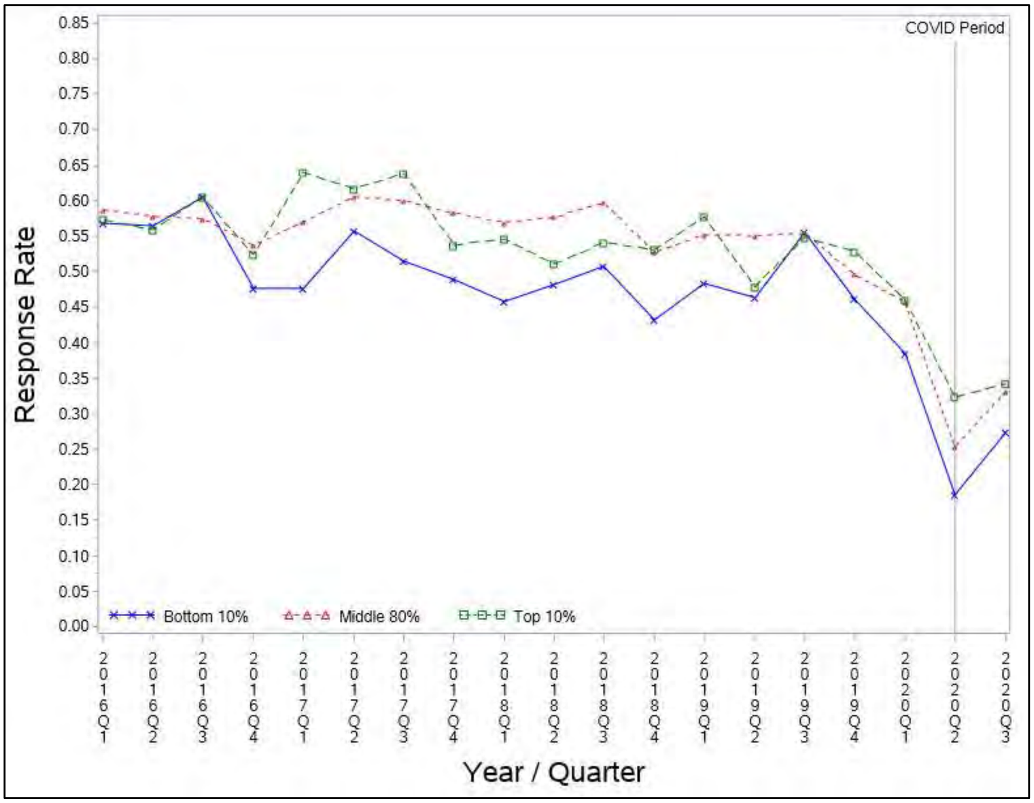


Figure 3 compares the response rates by the number of persons in the CU to see if there is a clear separation between their response rates. We see that the response rates for CUs with 1 person has been lower than the other number of persons from 2016Q1 to 2021Q1 with the exception of 2018Q3. Things change during the COVID-19 period: none of the response rates including the response rate for CUs with 1 person, show separation during the COVID-19 period. This suggests that the number of persons in the CU may not be as effective in the nonresponse adjustment for the COVID-19 period as it was in the past.

Figure 3 shows response rates by the three levels of ZIP code-level IRS income: top 10%, middle 80%, and bottom 10%. Every year the Internal Revenue Service (IRS) generates a dataset with summary-level information about the individual income tax returns filed in nearly every ZIP code of the United States. The datasets are publicly available on the IRS's website, and one of the pieces of information they contain is the average adjusted gross income per tax return by ZIP code. We use that information to stratify the ZIP codes into three categories: ZIP codes whose average adjusted gross income is in the top 10% of the distribution; ZIP codes whose average adjusted gross income is in the middle 80 percent of the distribution; and ZIP codes whose average adjusted gross income is in the bottom 10 percent of the distribution. These results are then merged with the CE data by the CUs' ZIP codes. Research by Sabelhaus et al. (2013) showed these ZIP-code income levels were a predictor of response rates for CE.

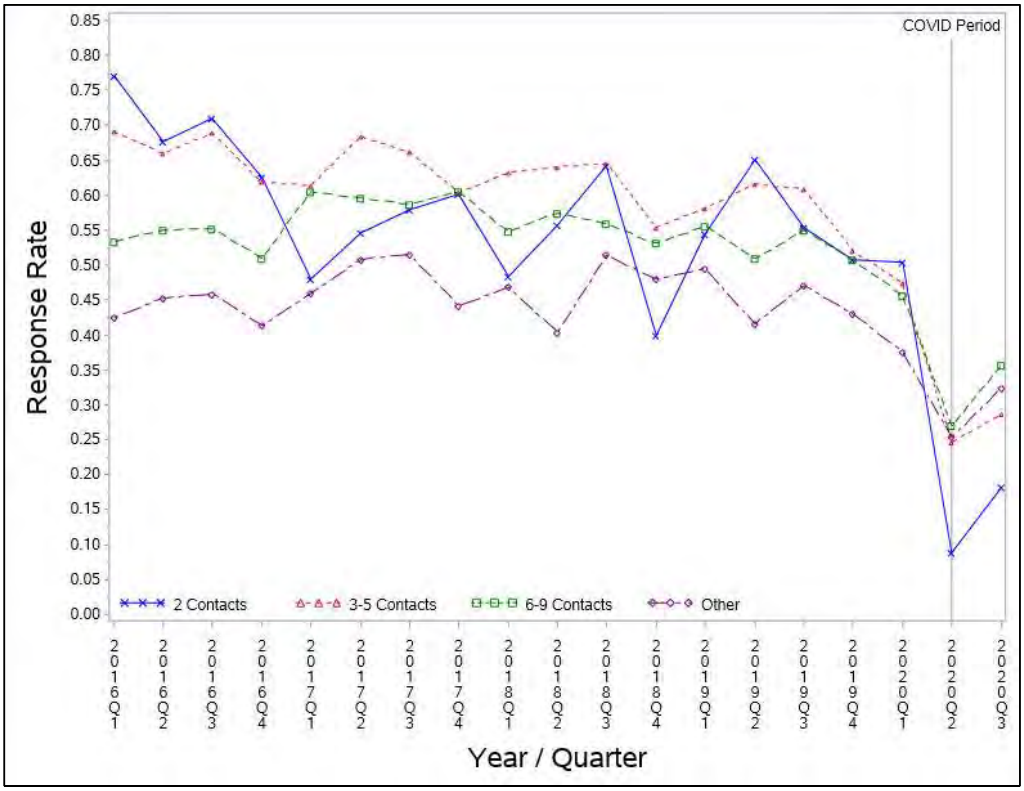
Figure 3. CE Diary Survey Response Rates by ZIP Code-Level IRS Income



In Figure 3, we see that sample CUs in ZIP codes with the Top 10 percent and Middle 80% of IRS income did not have much consistent separation from 2016Q1 to 2020Q3. However, from 2016Q4 to 2019Q1 and again from 2019Q4 to 2020Q3, the sample CUs in ZIP codes with the Bottom 10 percent of IRS income had the lowest response rates. This suggests that the IRS income variable may be a good predictor of response rates, including during the COVID-19 period.

Figure 4 shows response rates by the number of contact attempts during the interview process: 2, 3-5, 6-9, and other contact attempts.

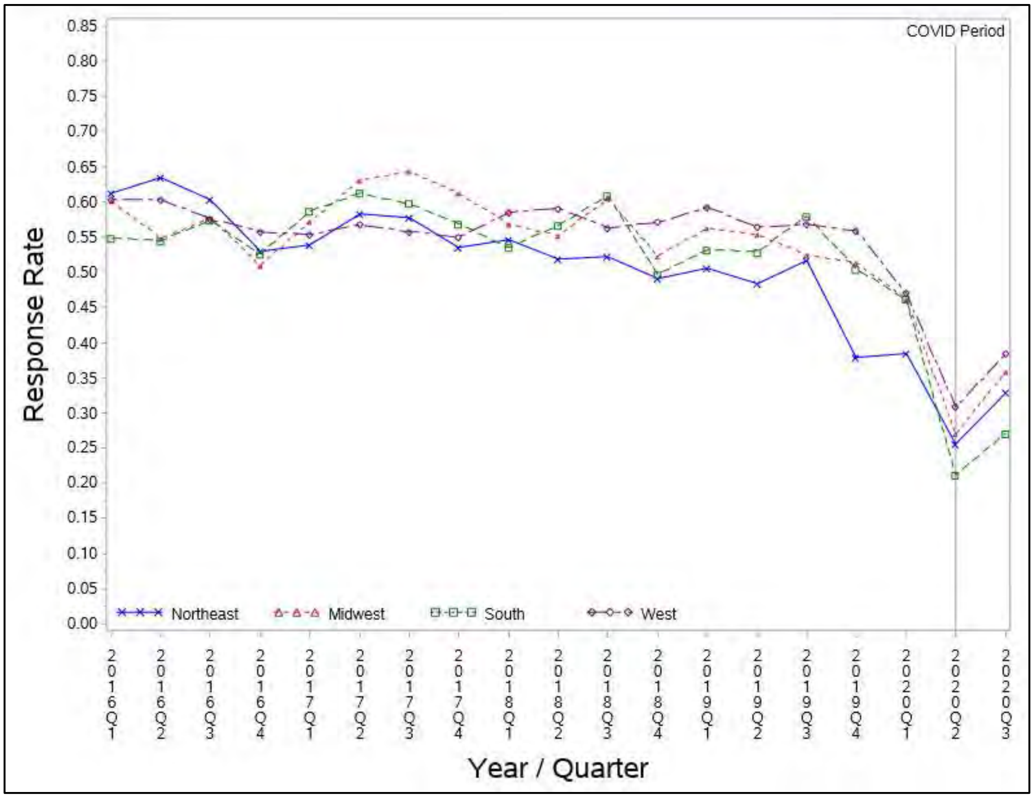
Figure 4. CE Diary Survey Response Rates by Number of Contacts



Although contact groups 3-5, 6-9, and other in Figure 4 exhibit a new trend and little separation during the COVID-19 period, the strong separation of contacts group 2 from the other groups suggests that the number of contacts has been, and still is, a good predictor of response rates during the COVID-19 period.

Figure 5 shows response rates by Census Regions: Northeast, South, Midwest, and West.

Figure 5: CE Diary Survey Response Rates by Census Regions



We see in Figure 5 that prior to the COVID-19 period, there has not been any consistent trend in any region having the highest or lowest response rates. However, during the COVID-19 period, the response rates for the South were lower than those for the other regions.

Conclusion for Analysis 1. The response rates for the variables used in the noninterview adjustment showed that some of the variables changed with respect to how they are associated with response. For example, the response rates for CUs with 2 contacts did not show any separation with the response rates for the other number of contact groups prior to the COVID-19 period but were lower during the COVID-19 period. We also saw that the response rate for 1 person in the CU was lower than the response rates for more than one person prior to the COVID-19 period but they were not different during the COVID-19 period. This exploratory evidence suggests that some of the past associations between response and the variables used in the noninterview adjustment have changed. However, it does not tell us whether the changes adversely affected the estimates of expenditures derived with the noninterview adjustments. The next section considers this question further.

Analysis 2: Comparison with Expenditure Estimates from Alternative Noninterview adjustments

To measure the bias in the estimates of expenditures due to nonresponse, we need to know the true value of expenditures for both respondents and nonrespondents. Of course, if we did know the true value of expenditures, we would not need to conduct the survey. The next best thing is to estimate expenditures using weights that have a noninterview adjustment tailored specifically to the COVID-19 period. We suggest that the estimates from the tailored noninterview adjustment have less bias than the current production noninterview adjustment and the differences in estimates produced by the two methods are a measure of the bias of the current estimates. This measure is not perfect because it measures the difference between the current weighting methodology and this tailored or "better" weighting methodology rather than between the current method and a "perfect" weighting methodology. However, this measure is an indication of possible improvement with respect to nonresponse bias. Rothbaum and Bee (2020) applied a similar approach to the Current Population Survey Annual Social and Economic Supplement.

To calculate the tailored noninterview adjustment for the COVID-19 period, we:

> **Examined several potential variables to see whether the noninterview adjustment process can be improved.** We want the best possible alternative noninterview adjustment for our comparisons. To do this, we considered all variables used by the current noninterview adjustment plus additional variables. Naturally, the number of variables we can consider is constrained since we need to know the value of the variables for both respondents and nonrespondents. **Error! Reference source not found.** lists all the variables we considered.[5]
>
> **Applied a different method for calculating the noninterview adjustment**. We used a logistic regression model to estimate the probability of a CU responding to the survey instead of the traditional cell method. The inverse of the estimated probability of response from this model

[5] Information about nonrespondents can be hard to obtain. Of the thirteen variables that were considered, three are geographic variables whose values are known for every address in the sample because of the inherent relationship between addresses and geography (REGION, DIVISION, SR_NSR); one is an economic variable whose values are obtained for every address in the sample by merging CE's database to a publicly available database from the Internal Revenue Service by the address's known ZIP code (WTIRSINC); six are demographic variables whose values are obtained for every address in the sample by merging CE's database to a publicly available database from the Census Bureau by the address's known tract number (Q1_65PLUS, Q1_Not_HS, Q1_BLW_POV, Q1_MED_INC, Q1_BLACK, Q1_HISP); one is an administrative variable whose values are obtained for every address in the sample by counting the number of contact attempts made by the survey's field representatives in the data collection process (WTNUMCNT); and two are demographic variables whose values are obtained for every nonrespondent in the sample by talking to their neighbors (WTMEMQ, WTTENURE). For a description of the variables considered, see Table 1.

was used as the noninterview adjustment. The logistic regression model allowed us to include more variables in the adjustment, and it also simplified the application by eliminating the need for certain operational steps like combining cells that do not have enough sample units in them.

Table 2 identifies the variables used in the current and alternative noninterview adjustments. The second column indicates whether the variable is used in the current noninterview adjustment, and the third and fourth columns indicate whether the variable was used in alternative noninterview adjustment (A2) or (A3).

Table 2: Variables of the Alternative Noninterview Adjustments

Variable	**Current Method**	**Alternative Methods (A2)**	**(A3)**
REGION	X		
DIVISION		X	X
WTIRSINC	X	X	
WTNUMCNT	X	X	X
MONTH			X
WTMEMQ	X		
WTTENURE		X	X
SR_NSR			
Q1_65PLUS			X
Q1_Not_HS			
Q1_BLW_POV			
Q1_MED_INC			
Q1_BLACK			
Q1_HISP		X	

X – indicates the variable was included in the given model.

Variables from the Census Planning Database. One way we were able to expand the potential list of variables for our noninterview adjustment is the use of tract-level information from the Census Planning Database (U.S. Census Bureau, 2015). We merged the tract-level estimates from the Planning Database to the CE Interview sample CUs. Then for each variable from the Planning Database, we found the quartiles of the variable using the sample CUs. The value of the quartiles was then assigned to the CUs of the sample for each variable.

The result is that each variable from the Planning Database defines a new variable with four values (1,2,3, and 4). These four values identify the four quartiles or the relative ranking of the CE sample tracts with respect to each variable. For example, a value of 1 for the variable Q1_HISP (proportional of tract's population that is Hispanic) means that the sample CU is in a tract that is in the bottom quartile of all CE sample tracts with respect to the proportion of the tract's Hispanic population. Therefore, CUs with a value of 1 have the smallest proportion of Hispanic population, and tracts with a value of 4 have the largest proportion of Hispanic population when considering the tracts of our sample.

The last seven variables of **Error! Reference source not found.**whose variable name starts with "Q1_" are quartiles of variables we derived from the Census Planning Database (U.S. Census Bureau, 2015). The variable name in square brackets in the description is the variable name on the Census Planning Database. Less than 2 percent of the Diary sample did not match the Census Planning Database. Since the percent of mismatches was so small, we included them in quartile 2 because we assumed that their average was similar to the overall average or one of the two middle quartiles.

An example of one of the quartile variables from the Census Planning Database is shown in Figure 6: quartiles of the percentages of the population that are Hispanic.

Figure 6. CE Diary Survey Response Rates by Quartiles of the Census Planning Database Estimate of the Percentage of Tract Hispanic

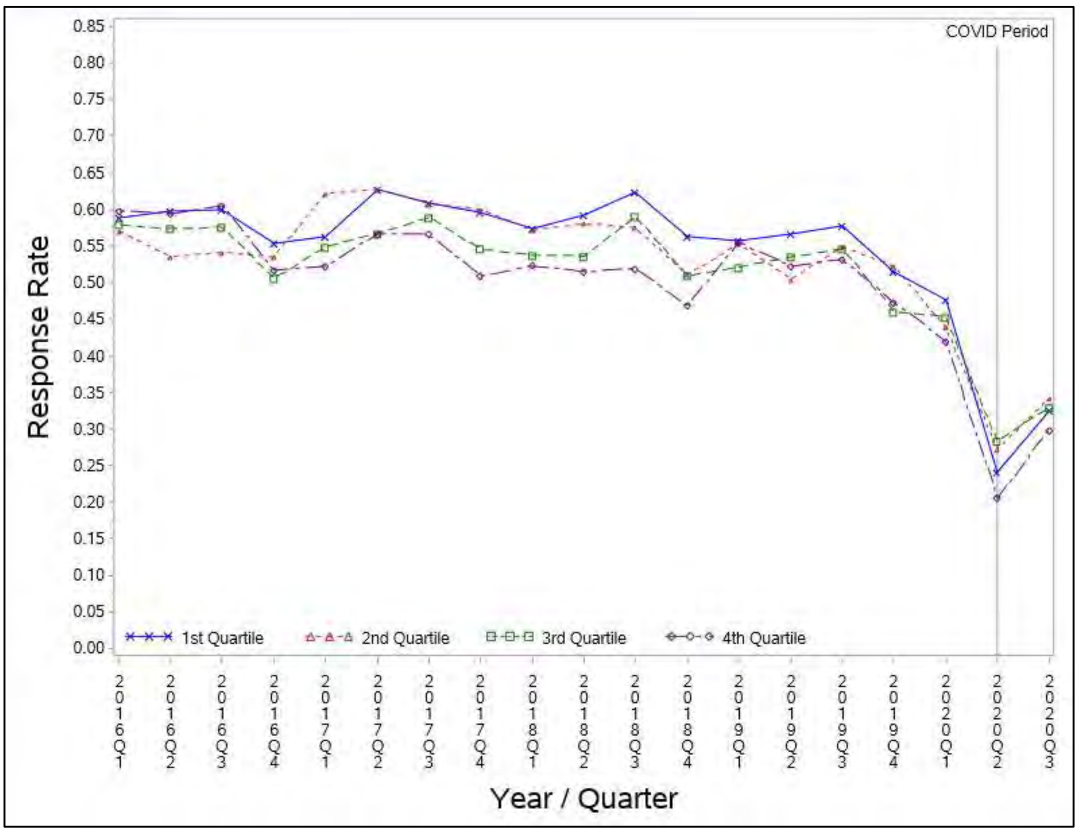


Figure 6 shows that response rates for the 1st and 4th quartiles have some separation with the 2nd and 3rd quartiles during the COVID-19 period. This suggests that the quartiles by Hispanic may be a good variable in modeling response rates during the COVID-19 period. As we will see in Analysis 2, it is a good variable for modeling response rates because we do include it in our 2020Q2 alternative noninterview adjustment later.

Selection of Variables for the Alternative Noninterview Adjustments

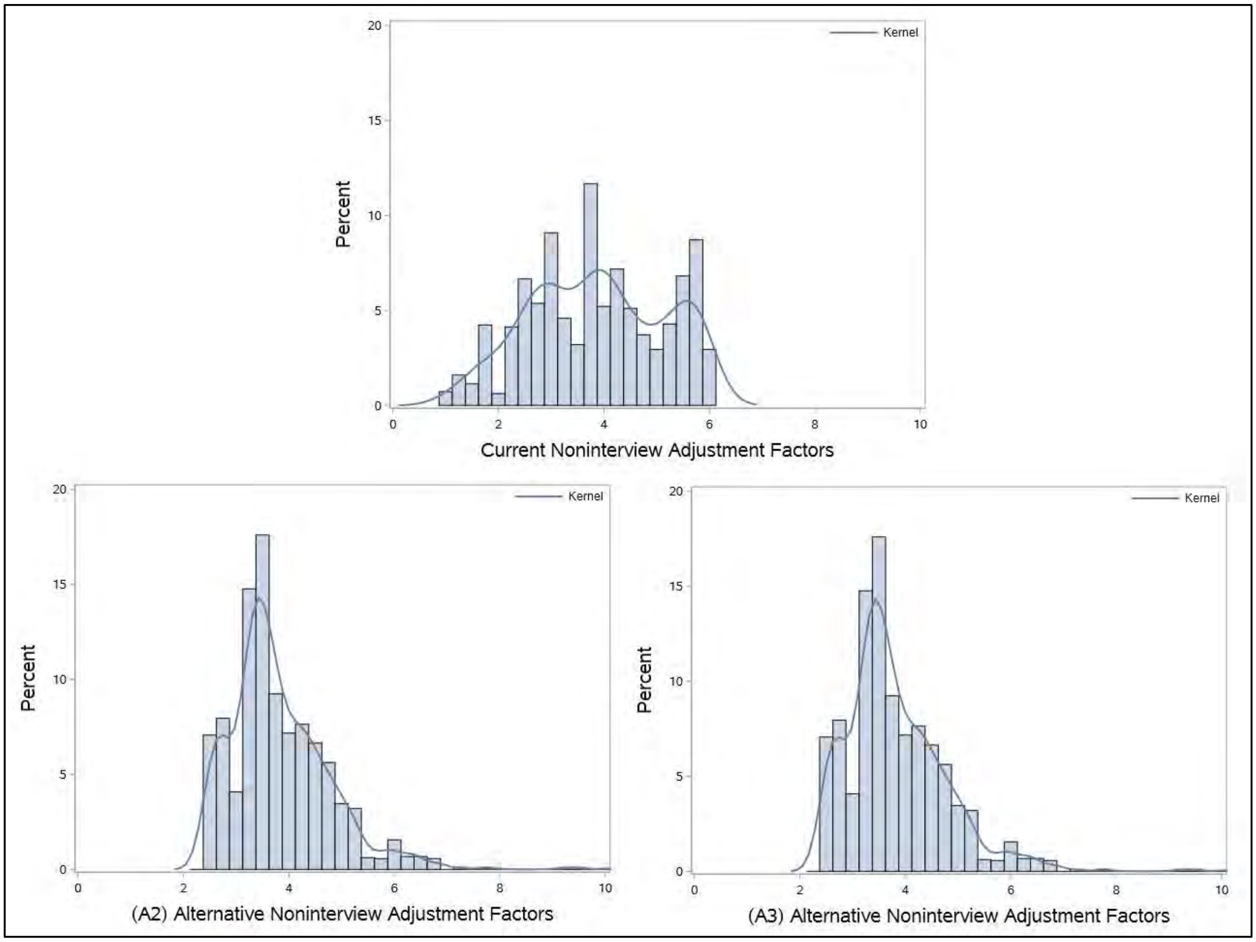
The last two columns of **Error! Reference source not found.** identify the two alternative models of the alternative noninterview adjustments. We refer to the models by the variables included in the model, which were all significant at a 5% level.

(A2): WTNUMCNT WTMEMQ WTTENURE DIVISION INT SR_NSR Q1_65PLUS Q1_Not_HS – This is our best model for predicting the 2020Q2 response rates.

(A3): WTNUMCNT WTMEMQ WTTENURE DIVISION INT SR_NSR – This is our best model for predicting the 2020Q3 response rates.

Using each of the alternatives models, we estimated the response rate for each respondent CU in 2020Q2. The inverse of each predicted response rate was used as the noninterview adjustment in the corresponding alternative weight. Figure 7 presents the distributions of the three noninterview adjustment factors as applied to the completed interviews of 2020Q2. The distribution on the top of Figure 7 is the distribution of the current noninterview adjustment factors; the distribution on the bottom-left side is the distribution of the alternative noninterview adjustment factors from model (A2); the distribution on the bottom-right side is the distribution of the alternative noninterview adjustment factors from model (A3).

Figure 7. Distributions of the 2020Q2 Alternative Noninterview Adjustment Factors



We provide the following two comments to assist the reading in interpreting Figure 7.

1) The current method caps the maximum value of the noninterview adjustment at 6, which accounts for the truncation of the distribution of the values of the current noninterview adjustment factor. We did not cap the noninterview adjustments for alternative methods 2 and 3 to minimize bias, even if that results in a higher variance. We allowed the alternative factor to be uncapped because we wanted to remove as much bias as possible.
2) The alternative noninterview adjustment factors have distributions that are similar to each other. However, the alternative noninterview adjustment factors (A2) and (A3) are different than the current noninterview adjustment, but not drastically different. This suggests that the alternative noninterview adjustment may have little impact on the estimates, which we consider further in Analysis 2.

Calibration of Weights. After the noninterview adjustment, the last step of the weighting process is the calibration of the weights to a set of control totals. Calibration has several purposes: to make the estimates agree with the control totals, reduce variances, and improve coverage in the case of under coverage. See Estevao and Särndal (2000), Kott (2012), and Särndal (2017) for more about the goals of calibration and how calibration achieves those goals. With the alternative methods, we calibrated the weights to the same known totals and used the same methodology as used with the current weights (Steinberg and Reyes-Morales 2017). See also Jayasuriya and Valliant (1995, 1996) for the general description of the methodology used by CE.

Results of Analysis 2: Comparison of Estimates of Mean Total Expenditures

This section compares the expenditure estimates produced by the current weights and the three alternative weights. The differences between the expenditure estimates under the current noninterview adjustment and the two alternative noninterview adjustments are a measure of nonresponse bias.

Table 3 presents the 2020Q2 national estimates of mean total expenditures for several large categories of expenditures.

Table 3: 2020Q2 Estimates from Alternative Noninterview Adjustments

	Current Method	Alternative Method (A2)		Alternative Method (A3)	
Mean of …	Estimate	Estimate	Percent Difference	Estimate	Percent Difference
Total Expenditures	35,390 (2,222)	35,834 (2,820)	1.3	36,638 (2,987)	3.5
Meals Purchased Away from Home Expenditures	1,551 (114)	1,582 (122)	2.0	1,603 (125)	3.4
Food Purchased for Home Consumption Expenditures	5,382 (299)	5,291 (246)	-1.7	5,347 (253)	-0.6
Clothing Expenditures	977 (148)	998 (138)	1.1	1,012 (137)	3.6
Other Expenditures	27,480 (2,099)	27,973 (2,699)	1.8	28,676 (2,851)	4.4

Note: Details may not sum to totals because of rounding.

The second column of Table 3 shows the estimates using the current noninterview adjustment with their standard errors in parentheses. The remaining columns show the results of the two alternative noninterview adjustments (A2) and (A3). For each alternative, we provide the estimate using the alternative noninterview adjustment and the percent difference between the current and the alternative.

Considering the method (A2) noninterview factors, which were developed just for 2020Q2, the largest magnitude of the difference is only 2.0 percent in Table 3. Considering both methods, the magnitude of the percent differences are all less than or equal to 4.4 percent, which suggests there is no evidence of nonresponse bias.

Table 4 presents the 2020Q3 national estimates of mean total expenditures for the same expenditure categories as used with 2020Q2.

Table 4: 2020Q3 Estimates from Alternative Noninterview Adjustments

Mean of …	Current Method	Alternative Method (A2)		Alternative Method (A3)	
	Estimate	Estimate	Percent Difference	Estimate	Percent Difference
Total Expenditures	45,070 (2,434)	43,674 (2,338)	-3.1	44,438 (2,408)	-1.5
Meals Purchased Away from Home Expenditures	2,233 (124)	2,197 (112)	-1.6	2,211 (114)	-1.0
Food Purchased for Home Consumption Expenditures	5,534 (190)	5,525 (191)	-0.2	5,597 (195)	1.1
Clothing Expenditures	1,371 (111)	1,377 (117)	0.4	1,380 (118)	0.6
Other Expenditures	35,932 (2,368)	34,576 (2,316)	-3.8	35,251 (2,356)	-1.9

Note: Details may not sum to totals because of rounding.

With Table 4, we see that magnitude of the 2020Q3 percent differences are all less than or equal to 2.0 percent for method (A3). Considering both methods, the magnitude of the percent differences are all less than or equal to 3.9 percent, which suggests there is no evidence of nonresponse bias. Therefore, we conclude that the estimates for 2020Q2 and 2020Q3 show no evidence of nonresponse bias.

Analysis 3: Comparison of Estimates Common to CE and ACS during the COVID-19 period

The goal of this analysis is to compare estimates common to the CE and ACS during the COVID-19 period. We use ACS as a benchmark because it is a well-known survey whose estimates are considered accurate. It also has small standard errors due to its large sample size; it has data on many socio-demographic characteristics, which makes the data analysis easier and makes its estimates methodologically consistent; and it produces annual estimates, which allows them to be studied as a time series. Similar comparisons of 2018 and 2019 are provided as context.

We define our relativity measure as

$$\text{Relativity of subgroup } i = \frac{\text{proportion of the population in a subgroup } i \text{ as estimated from CE}}{\text{proportion of the population in a subgroup } i \text{ as estimated from ACS}}$$

This relativity measure compares how well CE estimates agree with ACS estimates. It compares proportions of the relative measure of the size of different subgroups. Relatively values of close to 1.0mean that both CE and ACS produce the same estimate of the proportion for the given subgroup. This analysis does have an important limitation: agreement between CE and ACS could mean that both surveys had the same problem at the same time. We note this but suggest that it’s generally unlikely.

We use proportions as the totals from different surveys often do not agree; however, the proportion of a subgroup is a relative measure and should be reasonably similar in magnitude. For example, CE and ACS estimates of the total number of households by tenure in the U.S. may differ, however, both surveys should produce similar estimates of the proportion of households that are either owners or renters.

We use the fictitious example of Table 5 to further illustrate our measure of relativity.

Table 5: Fictitious Example of the Relativity Measure

	Proportion of Population		
	CE Estimate	ACS Estimate	Relativity
Homeowners	0.60	0.65	0.92
Renters	0.40	0.35	1.14
Overall	1.00	1.00	

For example, in a given year x, CE estimates that the proportion of homeowners and renters is 0.60 and 0.40, respectively. The ACS generates estimates of 0.65 and 0.35 respectively. The relativity measures of owners and renters would be 0.60/0.65 = 0.92 and 0.40/0.35 = 1.14, respectively. Assuming the ACS produces a more reliable estimate, this example would suggest that CE underestimates the homeowner population and overestimates the renter population.

Table 6 provides the CE estimates and the ACS source tables used in our comparisons.

Table 6: Relativity Estimates and their Source Tables for Comparable Estimates from the American Community Survey

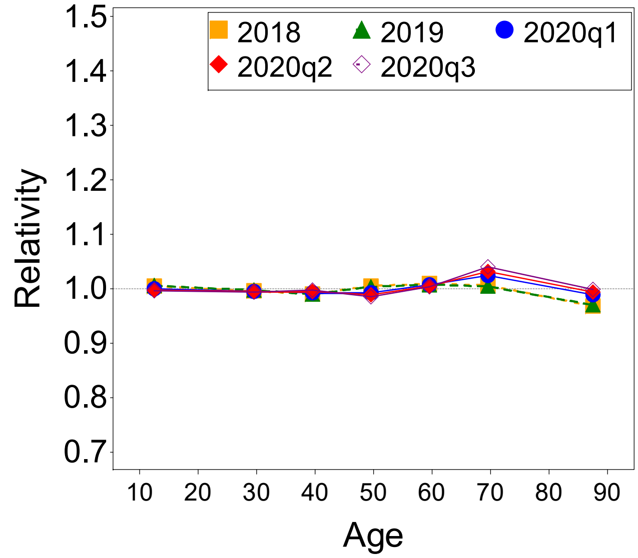
CE Estimate of…	Source Table for Comparable ACS Estimate
Population by age	SEX BY AGE (B01001)
Consumer Units by tenure	TENURE BY HOUSEHOLD SIZE (B25009)
Consumer Units by number of persons in the Consumer Unit	TENURE BY HOUSEHOLD SIZE (B25009)
Population by level of education	SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER (B15002)
Consumer Units by income	HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2019 INFLATION-ADJUSTED DOLLARS)[6] (B19001)

The value in parentheses in the second column of Table 6 identifies the specific ACS Table that we used as the source for the ACS estimates. We compared the 2018 and 2019 CE estimates with the 2018 and 2019 1-year ACS estimates, respectively. At the time, the 2020 1-year ACS estimates were not available, we compared the 2020Q2 estimates with the 2019 1-year ACS estimates. Such a comparison assumes the proportions stay relatively constant from year-to-year.

Figure 8 shows the relativity measures for age.

[6] Household incomes are inflation-adjusted in the American Community Survey with the Consumer Price Index (CPI). Specifically, ACS uses annual averages of the national All-Items CPI-U-RS (CPI Research Series) index to adjust the household income estimates.

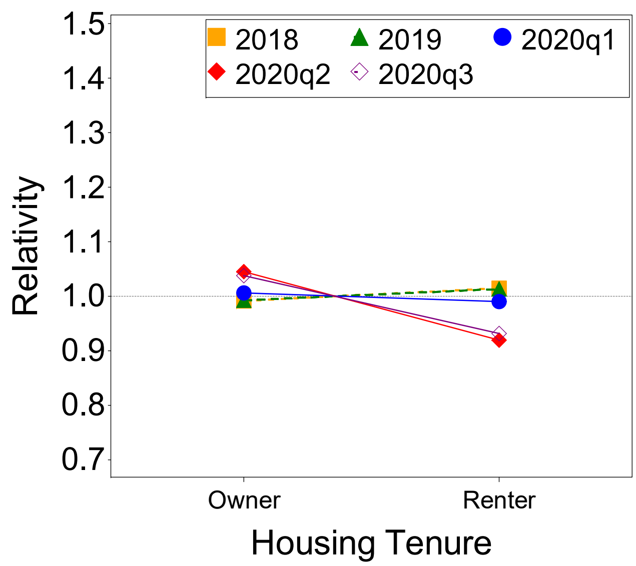
Figure 8: Relativity Measures by Age



The relativity measures in Figure 8 for age are very close to 1.0 which means the CE and ACS estimates of the percent of the population by age are in close agreement. This is expected since CE calibrates to population totals by age. The relativity measures are close to 1.0 for the COVID-19 period and the prior periods, 2018 and 2019.

Figure 9 shows the relativity measures for tenure.

Figure 9: Relativity Measures by Housing Tenure



The relativity measures in Figure 9 for housing tenure are very close to 1.0 for 2018 and 2019 showing close agreement between the CE and ACS estimates. However, there are differences between 2020Q2 and 2020Q3. Beginning with the COVID period in 2020Q2, the CPS had lower than normal response rates and higher than normal rates of homeownership than in previous quarters (Zhao 2022). The strong correlation between the change in the response rate and the change in the home ownership rate suggests that some or all of the differences are due to the CPS response rates (U.S. Census Bureau 2020). The next section discusses the impact of the Current Population Survey / Housing Vacancy Survey (CPS/HVS) response rates on CE's tenure estimates.

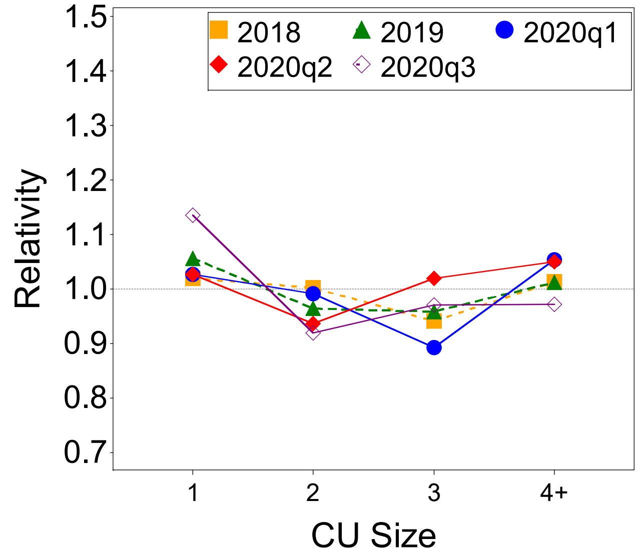
Impact of CPS's Lower Response Rates due to COVID-19. The COVID-19 pandemic affected data collection operations not just for the CE survey, but also for CPS at the end of 2020Q1 and especially 2020Q2. The suspension of in-person interviews affected the entirety of the data collection period for the second quarter of April, May, and June. Like CE, the CPS weights adjust for changes over time in the overall response rate by adjusting the weights to ensure that the estimates total to the overall number of housing units in the United States and several other control totals. However, their weighting methodology does not adjust for differences in response rates between homeowners and renters.

The CPS/HVS response rate was slightly above 80-percent in 2019, then it decreased to 79 percent in 2020Q1, and fell to 67 percent in 2020Q2. It is not clear whether CPS's decreasing response rates were correlated with households' tenure (owner/renter) status. However, the homeownership rate has traditionally been nearly 10 percentage points lower for in-person interviews than for other modes of interview, so the suspension of in-person interviews is the likely reason CPS/HVS's homeownership rate suddenly increased in 2020Q2. CPS/HVS lower response rates for homeownership directly affected CE's control totals used in calibration. This was detected when CE generated its own 2020Q2 population counts and noticed that there was an unexpected increase in the number of owner households by more than 3 million, or nearly 4 percent from the prior quarter. Therefore, the relativity measures for tenure in 2020Q2 and 2020Q3 are most likely impacted by the CPS response rates. Using CPS/HVS's 2020Q3 data, it still shows a higher homeownership rate than the 2021Q1 and prior, but the September data are closer to the pre COVID period.

Therefore, this should be considered when data users make inferences pertaining to CE's tenure variable when using 2020Q2 data. In general, homeowners have higher expenditures than renters, so if the homeowners' weights are overstated relative to renters, then the resulting expenditures may be overstated at the national level because of this.

Figure 10 shows the relativity measures for size of the CU.

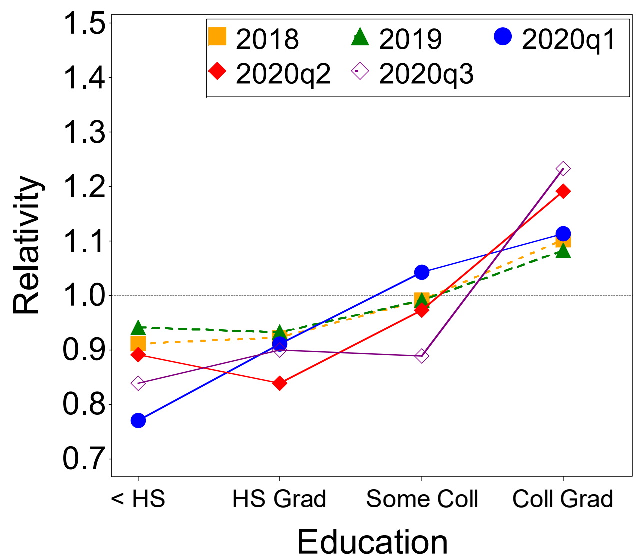
Figure 10: Relativity Measures by Size of Consumer Unit



The relativity measures in Figure 10 are not as close to 1.0 as the relativity measures for age in Figure 9. This is most likely due to calibration: the CU Size variable is not used in calibration but is used in noninterview adjustment. During the third quarter of 2020, it appears that the relativity measure for CUs with 1 person is further from 1.00 than in the periods prior to COVID-19. Conversely, the relativity measures for that quarter were lower for the CUs with larger size.

Figure 11 shows the relativity measures for the level of education.

Figure 11: Relativity Measures by Level of Education



The relativity measures by level of education in Figure 11 show that CE and ACS are not always in agreement, but we do see that the three quarters of data from 2020 show a noticeably lower relativity measure for those CUs with less than a high school education when compared to 2018 and 2019. Conversely, the three quarters of data from 2020 show a noticeably higher relativity measure for college graduates than in the prior years.

Figure 12 shows the relativity measures by the income of the CU.

Figure 12: Relativity Measures by Income

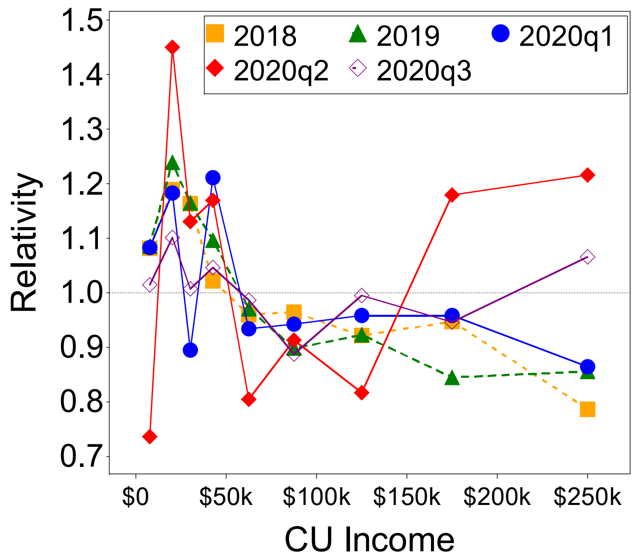


Figure 12 shows that the relativities during the second quarter of 2020, at the height of the COVID-19 pandemic, show that the lowest income group was noticeably underrepresented when compared to the ACS, and to other time periods. Conversely, the two higher income groups had higher relativity measures than other time periods.

However, the third quarter of 2020 shows relativity measures closer to 1.00 for all income groups. This implies that the data distribution is much more consistent with that of the ACS.

Conclusion of Analysis 3.
Other than the aforementioned example of housing tenure, the most noticeable effects of the COVID-19 period on relativities pertained to the variables of Education and Income. CUs with both lower education levels and income levels were less prevalent during the COVID-19 period; conversely, CUs with both higher education levels and income levels were more prevalent during the COVID-19 period. As income and education levels are highly correlated, the similarity of these effects is not surprising.

Conclusion

We found no evidence of nonresponse bias in the expenditure estimates related to the COVID-19 pandemic. Analysis 1 showed that the response rates for the different variables used in the noninterview adjustment showed that some of the variables changed with respect to how they are associated with response. In Analysis 2, we produced expenditure estimates with three alternative weighting procedures that were tailored to the COVID-19 period, and all three expenditure estimates were in reasonable agreement with those computed with the normal weighting procedure used to produce public estimates. In Analysis 3, we compared estimates common to both CE with ACS and found that the CE estimates during the COVID-19 period were not greatly different than the ACS estimates. This evidence suggests that CE was not impacted by nonresponse bias during the COVID-19 pandemic.

References

Armstrong, G., Grayson, G., Miller, T. and Pham, S. (2022). "Reference Guide CE Data Quality Profile," Consumer Expenditure Surveys Program Report Series, April 29, 2021.
https://www.bls.gov/cex/research_papers/pdf/dqp-reference-guide-midyear-2021.pdf

Ash, S., Nix, B., and Steinberg, B. (2022). U.S. Bureau of Labor Statistics Report. "Report on Nonresponse Bias during the COVID-19 Period for the Consumer Expenditures Interview Survey," dated March 10, 2022.
https://www.bls.gov/cex/research_papers/pdf/ash-nix-steinberg-covid-nonresponse-diary-report.pdf

Bethlehem, J., Cobben, F., and Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*, John Wiley & Sons.

Bethlehem, J., and Kersten, H.M.P. (1985). "On the Treatment of Nonresponse in Sample Surveys," Journal of Official Statistics, 287-300.

Brick, J.M. (2013). "Unit Nonresponse and Weighting Adjustments: A Critical Review," Journal of Official Statistics, 29, 329–353.

Chopova, B., Edgar, J., Gonzalez, J., King, S., McGrath, D.E. and Tan, L. (2008). U.S. Bureau of Labor Statistics Report. "Assessing Nonresponse Bias in the CE Interview survey," dated 2008.
https://www.bls.gov/osmr/research-papers/2009/st090220.htm

Dumbacher, B., Swanson, D., Reyes-Morales, S., Steinberg, B., Ash, S., and Castelo, D. (2012). U.S. Census Bureau memorandum entitled "Consumer Expenditure Surveys 2010 Redesign Research: Noninterview and Calibration Weighting Adjustments for the Consumer Expenditure Surveys," from Ruth Ann Killion to Jay Ryan, dated September 26, 2012.

Estevao, V.M. and Särndal, C.-E. (2000). A Functional Form Approach to Calibration, Journal of Official Statistics, 16, 4, 379-399.

Jayasuriya, B.R. and Valliant, R. (1995). An Application of Regression and Calibration Estimation to Post-Stratification in a Household Survey, Joint Statistical Meetings, Section on Survey Research Methods.

Jayasuriya, B.R. and Valliant, R. (1996). An application of Restricted Regression Estimation in Household Surveys, Survey Methodology, 22, 127-137.

Knappenberger, C., Lee, Y., Pham, S., and Armstrong, G. (2021). “The Midyear CE Data Quality Profile - 2020,” Consumer Expenditure Surveys Program Report Series, April 29, 2021. https://www.bls.gov/cex/research_papers/pdf/dqreport-midyear-2020.pdf

Kott, P.S. (2012). An Introduction to Calibration Weighting for Establishment Surveys, Proceedings of the Fourth International Conference on Establishment Surveys.

Lee, Y., and Biagas, D. (2021). “COVID-19 Response and Data Quality Impacts in the CE Interview Survey,” Consumer Expenditure Surveys Program Report Series, October 2021. https://www.bls.gov/cex/research_papers/pdf/lee-biagas-covid-19-response-and-data-quality-impacts-in-the-ce-interview-survey.pdf

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review, 54, 139–157.

McBride, B., Graf, N., Edwards, M. and Smith, M.(2021). “COVID-19 CE Diary Preliminary Report: Response and Data Quality Impacts,” Consumer Expenditure Surveys Program Report Series, June, 2021. https://www.bls.gov/cex/research_papers/pdf/mcbride-et-al-COVID-CE-diary-preliminary-report-response-and-data-quality-impacts.pdf

Oh, H.L. and Scheuren, F.S. (1983). “Weighting adjustment for unit nonresponse.” In Incomplete Data in Sample Surveys, 2, Ed. W.G. Madow, I. Olkin and D. B. Rubin, pp. 143-184. New York: Academic Press.

Rothbaum, J. and Bee, A. (2020). “Coronavirus Infects Surveys, Too: Nonresponse Bias During the Pandemic in the CPS ASEC,” SEHSD Working Paper Number 2020-10, September 15, 2020. https://www.census.gov/library/working-papers/2020/demo/SEHSD-WP2020-10.html

Sabelhaus J., Johnson, D., Ash, S., Swanson, D., Garner, T. and Henderson, S. (2013). “*Is the Consumer Expenditure Survey Representative by Income?*,” (NBER Working Paper No. 19589, October 2013). https://www.nber.org/system/files/working_papers/w19589/w19589.pdf

Särndal, C.-E. (2007). “The calibration approach in survey theory and practice,” Survey Methodology, 33, 2, 99-1.

Särndal, C.-E., and Swensson, B. (1987). “A General View of Estimation for Two Phases of Selection with Applications to Two Phase Sampling and Nonresponse,” International Statistical Review, 55, 3, 279-294.

Steinberg, B., Krieger, S., McBride, B., Nix, B., Sverchkov, M. and Yang, D. (2022). U.S. Bureau of Labor Statistics Report “A Nonresponse Bias Study of the Consumer Expenditure Survey for the Ten-Year Period 2010-2019.” https://www.bls.gov/cex/research_papers/pdf/steinberg-et-al-a-nonresponse-bias-study-of-the-consumer-expenditure-survey-for-the-ten-year-period-2010-2019.pdf

Steinberg, B. and Reyes-Morales, S. E. (2017). Bureau of Labor Statistics memorandum “Specifications for Weighting the Interview and Diary Surveys Using the Calibration Method – 2018,” dated November 17, 2017.

U.S. Bureau of Labor Statistics (2018). “Handbook of Methods,” last updated September 12, 2022. https://www.bls.gov/opub/hom/cex/home.htm

U.S. Census Bureau (2015). “Planning Database with 2010 Census and 2009 – 2013 American Community Survey Data, At the Tract Level,” July 25, 2015. https://www.census.gov/data/developers/data-sets/planning-database.2015.html

U.S. Census Bureau (2020). “Frequently asked questions: The impact of the coronavirus (COVID-19) pandemic on the Current Population Survey/Housing Vacancy Survey (CPS/HVS).” https://www.census.gov/housing/hvs/files/qtr220/impact_coronavirus_20q2.pdf

Vartivarian, S. and Little R.A. (2002). “On the Formation of Weighting Adjustment Cells for Nonresponse Adjustment,” Proceedings of the Section on Survey Research Methods, 3553-3558.

Zhao, N. (2022). “Homeownership Rate Stable at 65.4%,” National Association of Home Builders, dated April 27 (accessed August 5, 2022). https://eyeonhousing.org/2022/04/homeownership-rate-stable-at-65-4/

Consumer Expenditures Diary Survey Response Rates by Selected Characteristics

Figure A1: CE Diary Survey Response Rates by Tenure Status

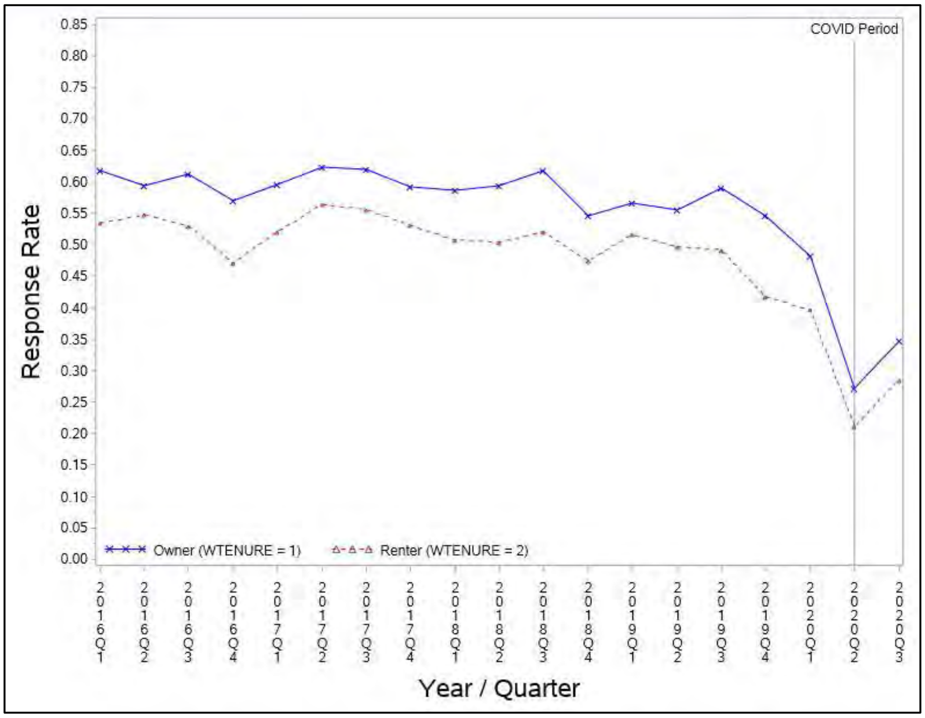
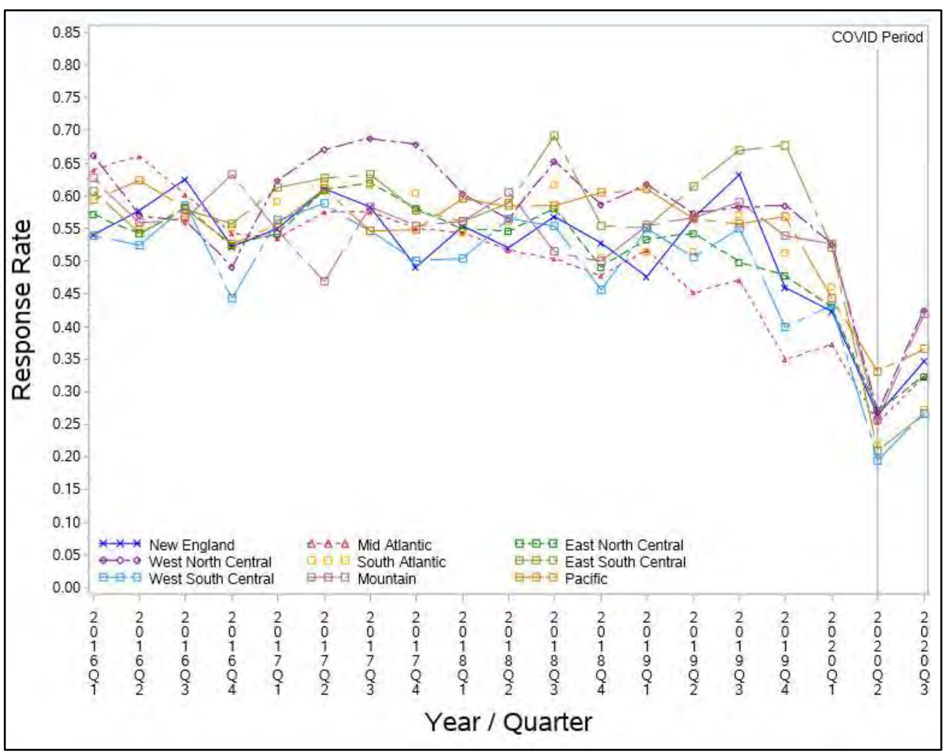


Figure A2: Response Rates by Census Division



Consumer Expenditures Diary Survey Response Rates by Selected Characteristics

Figure A3: Response Rates by Urban/Rural

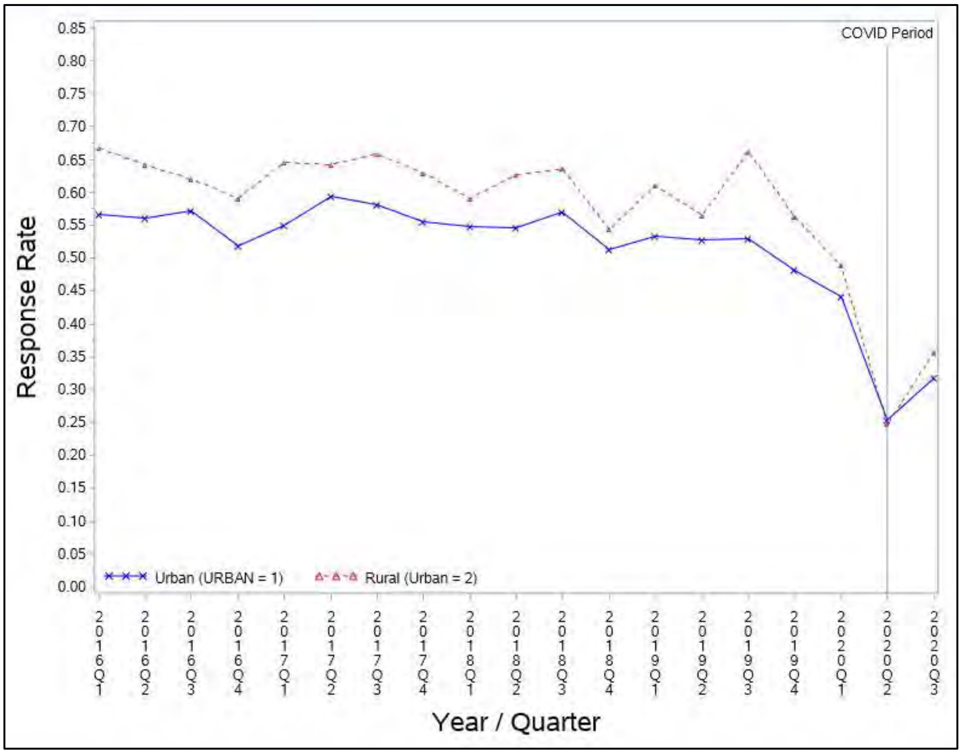
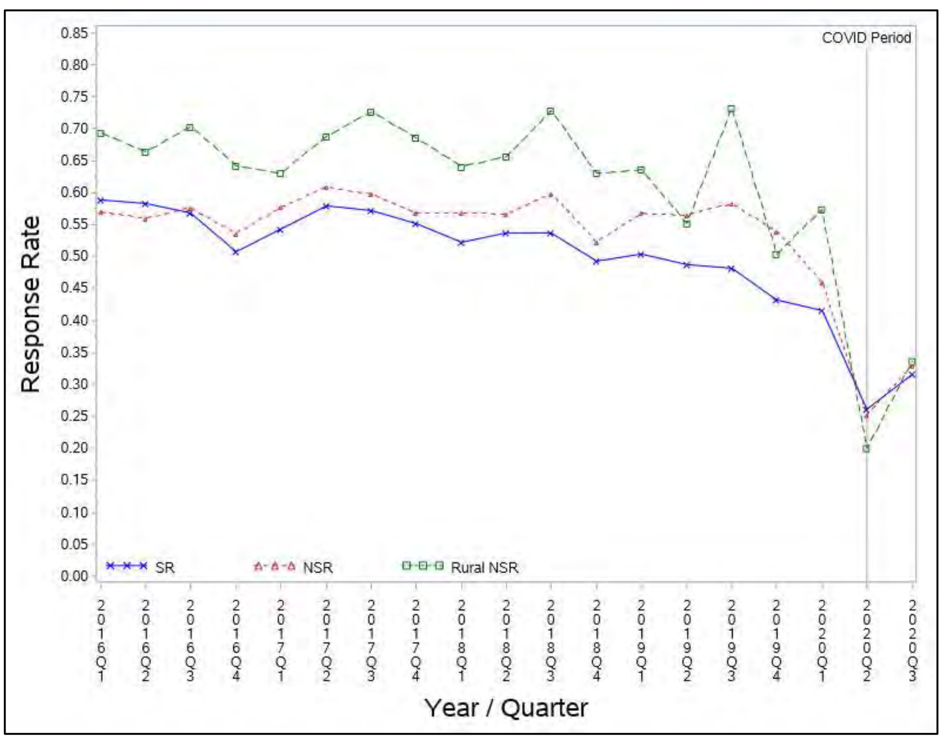


Figure A4: Response Rates by Size of Strata



Consumer Expenditures Diary Survey Response Rates by Selected Characteristics

Figure A5: Response Rates by Quartiles of Percent Not High School Graduate Population

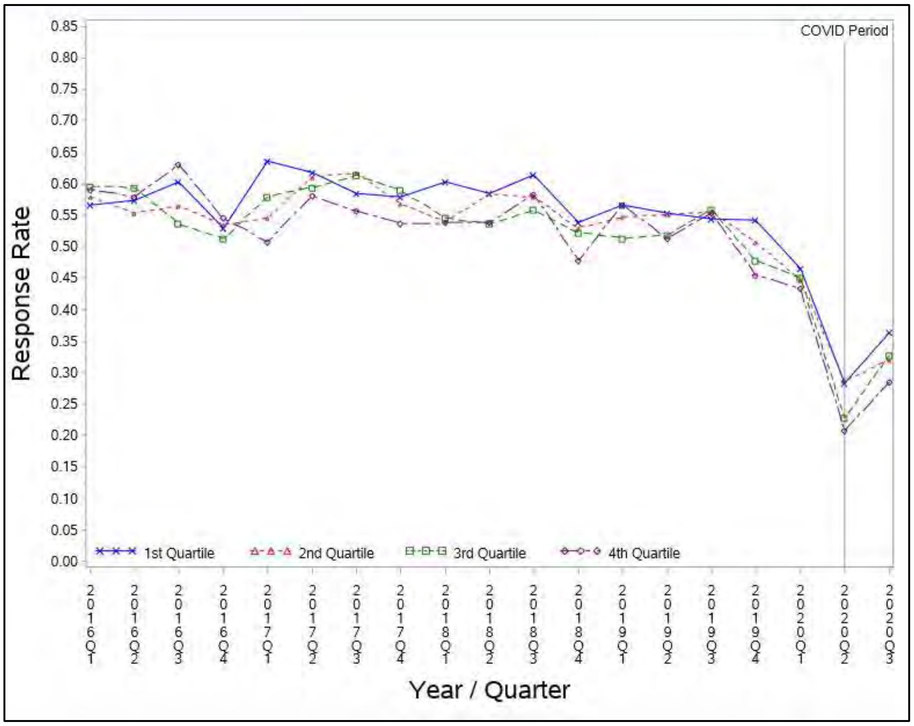
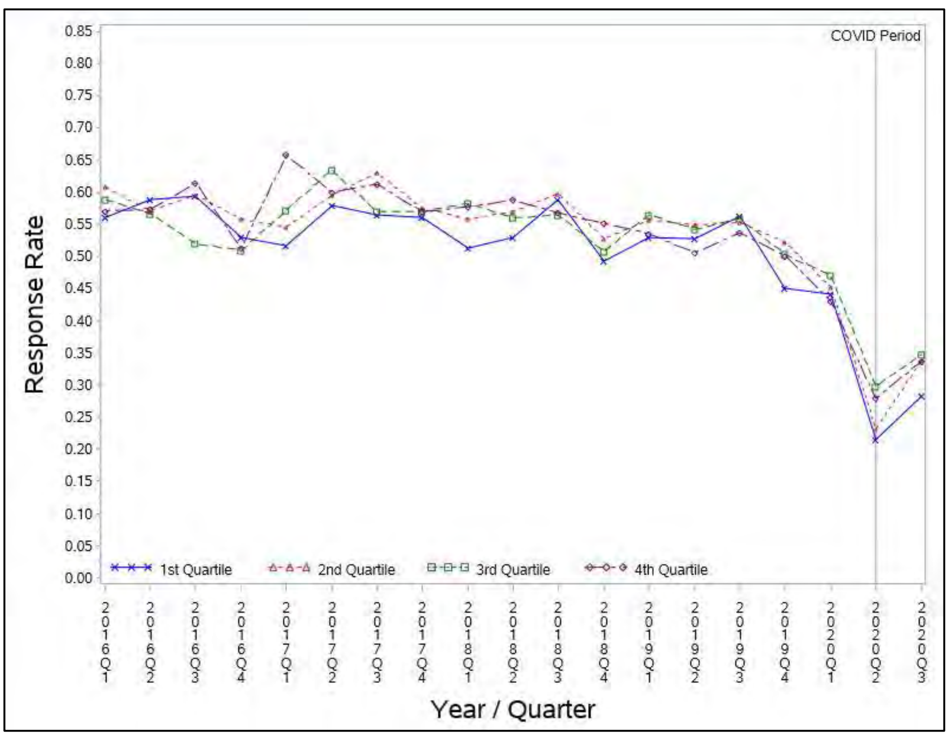


Figure A6: Response Rates by Quartiles of Median Income



Consumer Expenditures Diary Survey Response Rates by Selected Characteristics

Figure A7: Response Rates by Quartiles of Percent Black Alone Population

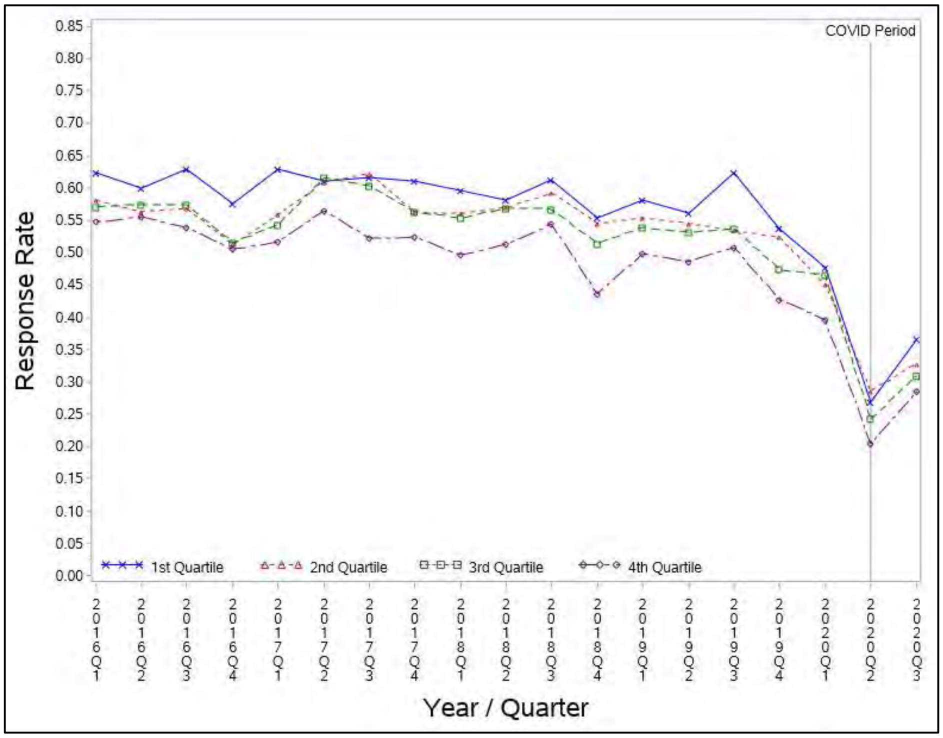
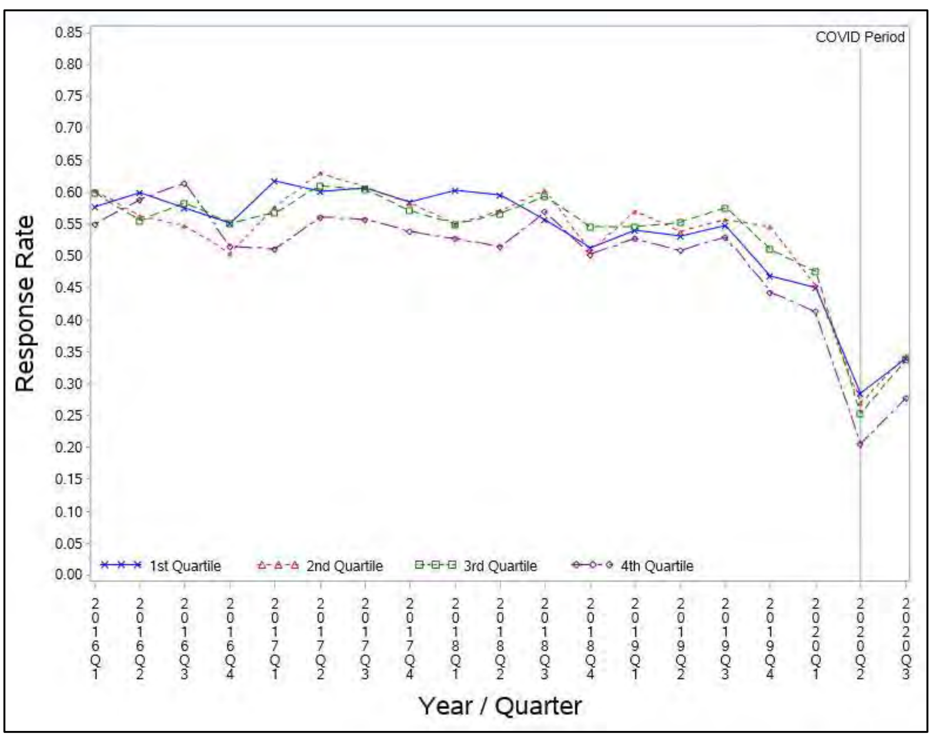


Figure A8: Response Rates by Quartiles of Percent Households under Poverty Level



Consumer Expenditures Diary Survey Response Rates by Selected Characteristics

Figure A9: Response Rates by Quartiles of Percent Aged 65+