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

March 10, 2022

Stephen Ash, Brian Nix, and Barry Steinberg

*Consumer Expenditure Surveys Program Report Series*

**Overview**

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

**Background**

The emergence of COVID-19 in March 2020 has caused extraordinary changes to many facets of life, including the Consumer Expenditures Surveys (CE). For example, as a result of COVID-19, the Bureau of Labor Statistics enacted new interviewing protocols that impacted response rates (Knappenberger, Lee, Pham, and Armstrong 2021a)

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. The "maximum telephone interviewing" protocol began on March 19, 2020, and it was gradually phased out over the summer and fall of 2020. Most of the survey's data in most of the country were collected by telephone during the second and third quarters of 2020 (Knappenberger, Lee, Pham, and Armstrong 2021a, 2021b).

Figure 1 presents overall response rates for the Interview survey.

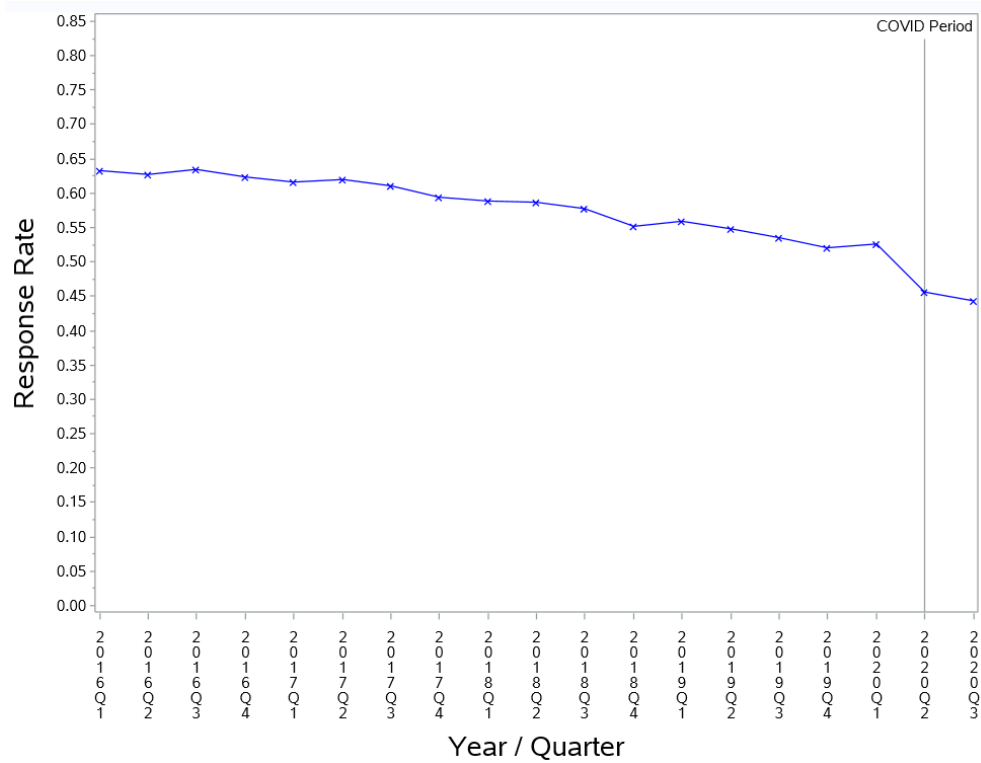**Figure 1: Response Rates for Household Surveys**



Figure 1 shows that the response rates have decreased over time, but the decrease in the COVID-19 period (2020Q2 to 2020Q3) was larger than in the past. Two studies have been released examining the impact of COVID-19 on response rates and data quality in the Consumer Expenditure Surveys, one for the Interview Survey (Lee and Biagas, 2021) and one for the Diary Survey (McBride and Graf, 2021). This report goes one step further and examines the question of whether the decrease in response rates during the COVID-19 period generated additional nonresponse bias due to the COVID-19 pandemic. This study is for the Interview Survey, and a separate study of the Diary Survey is being conducted.

Prior comprehensive studies of nonresponse bias for the CE surveys include Chopova et al. (2008) and Steinberg et al. (2020). For more background on the CE surveys, see the *Handbook of Methods* (U.S. Bureau of Labor Statistics 2018).

The remainder of the report is organized as follows. The next section defines nonresponse bias with respect to estimates and explains how the nonresponse adjustment in the weighting is used to reduce nonresponse bias. We also provide a summary of the nonresponse adjustment method currently used by CE weights. The following three sections provide separate analyses of nonresponse and nonresponse bias.
- 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 from the COVID-19 period with estimates from the American Community Survey (ACS). Similar comparisons of 2018 and 2019 estimates with ACS estimates are provided as context.

**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.[1] The U.S. Census Bureau[2] selects a representative sample of addresses for the survey, and then its field representatives visit those addresses and attempt to collect expenditure data from the CUs that live there. Some of the addresses are *ineligible* for the survey because they do not have CUs living there, such as vacant housing units, businesses, or other nonresidential addresses. However, for addresses having 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 nonresponse adjustment process, and then those nonresponse-adjusted weights are adjusted to the number of "known" population totals from the Current Population Survey (CPS) in a calibration adjustment process. This report focuses on the nonresponse adjustment process.

The nonresponse adjustment process uses the cell adjustment method, where 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. 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).

---

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

[2] The Bureau of Labor Statistics works with the U.S. 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 et al. 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, and 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, nonresponse 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.

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 less biased the survey's estimates will be with respect to both nonresponse and the variable of interest, and therefore the better the estimates are likely to be.

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. Currently, the variables used to define the cells in terms of nonresponse include: WTMEMQ (number of persons in the CU); WTIRSINC (IRS income); WTNUMCNTS (number of contact attempts); and REGION (Census Region). These variables are defined more fully in subsequent Table 1.

The next three sections describe the three different sets of analyses that we conducted.

**Analysis 1: Exploratory Analysis of Response Rates**

In this analysis, we review the effectiveness of the variables currently used in the nonresponse adjustment process. 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, while ineffective 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

5

effective the variable, the farther apart the lines in the graphs. In this analysis we are looking at the effectiveness of the variables and whether their effectiveness changed in the COVID-19 period. The variables currently are: WTMEMQ (number of persons in the CU); WTIRSINC (IRS income); WTNUMCNTS (number of contact attempts); and REGION (Census Region). These variables are defined more fully in subsequent Table 1.

Although there are no formal statistical tests associated with the graphs in this section, we think they provide useful background for an understanding of the response rates. The appendix includes similar graphs of response rates for the other variables considered in our analysis.

Figure 2 shows response rates by the number of persons in the CU: 1, 2, 3 to 4, and 5+ persons.

**Figure 2: CE Interview Survey Response Rates by Number of Persons in the Consumer Unit**
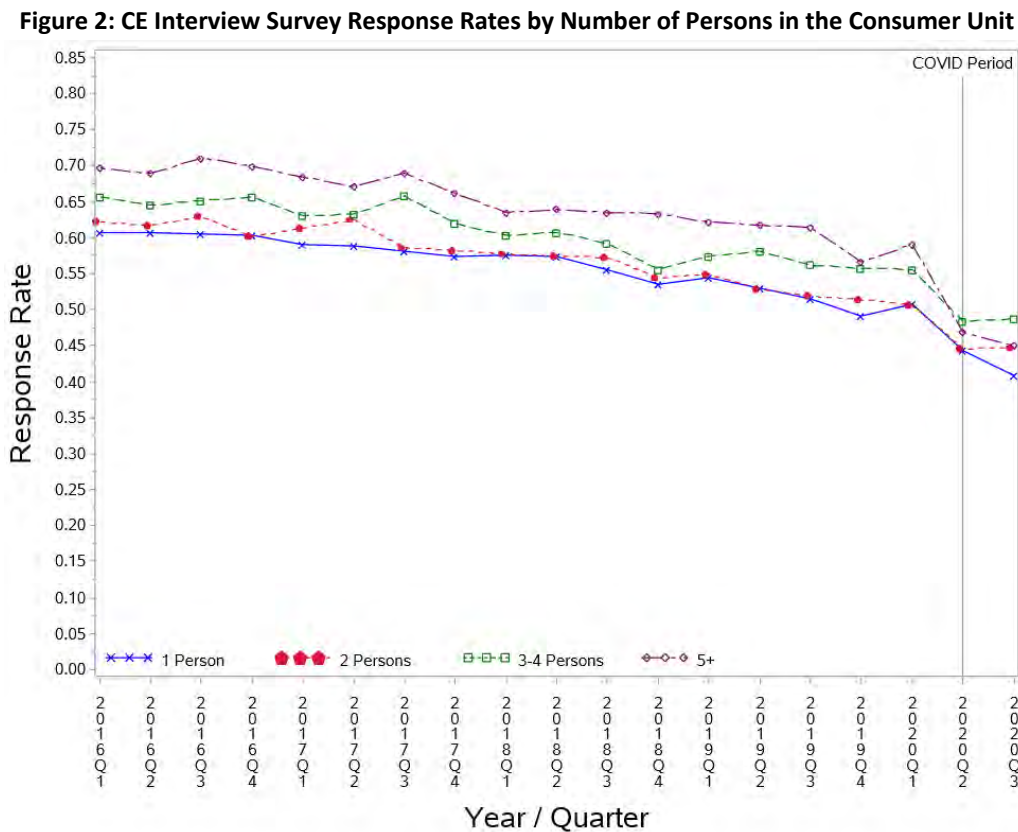


Figure 2 shows that CUs composed of one member are the least likely to respond to the survey, and CUs with 5 or more persons are most likely to respond to the survey.

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. (2015) showed these zip code level incomes from the IRS were a predictor of response rates for CE.

**Figure 3. CE Interview Survey Response Rates by Zip code-level IRS Income**
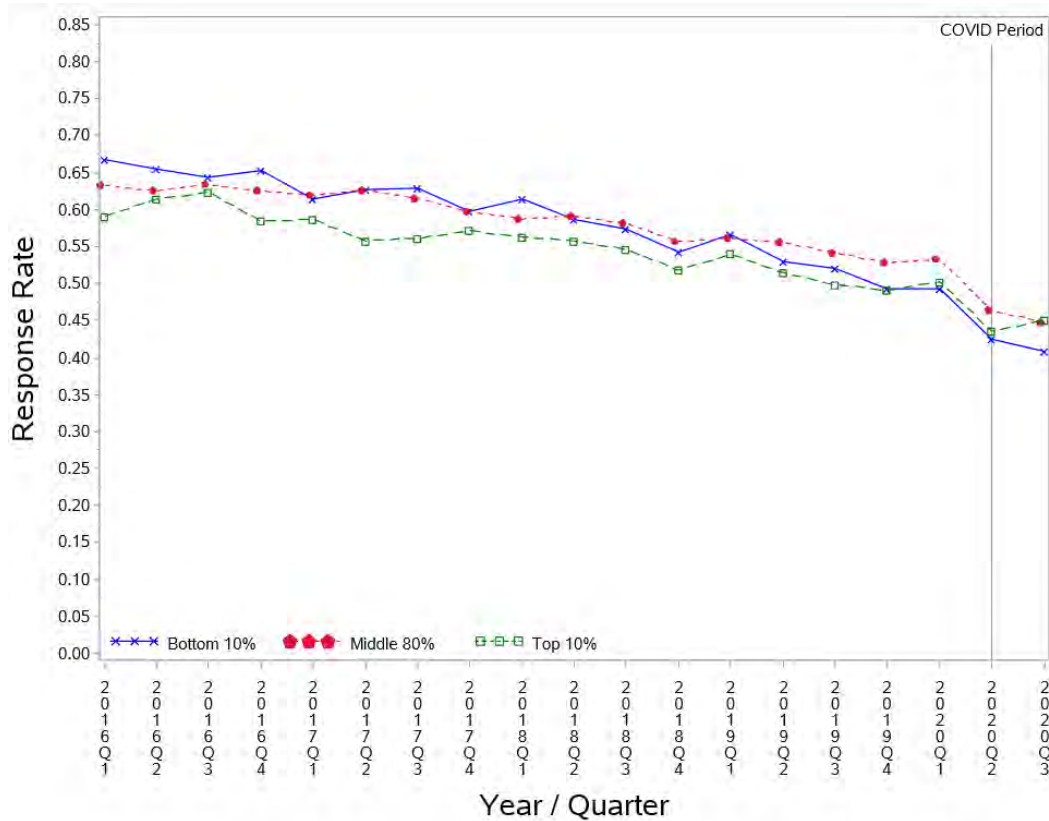


In Figure 2, we see that the different groups formed by IRS Income show different amounts of separation at different times. Prior to 2019Q4, the sample CUs in zip codes with the Top 10 percent of IRS income distinctly had the lowest response rates. However, from 2019Q4 through 2020Q2 there does not appear to be much difference between the bottom 10 percent and the top 10 percent, while in 2020Q3 there does not appear to be much difference between the top 10 percent and the middle 80 percent. This suggests that the IRS income variable may not be a strong predictor of response rates in the COVID-19 period. However, given the change in pattern starts before the COVID-19 period, and there are few data points available during the COVID-19 period as of the time of this study, whether this is a COVID-19 effect or not is uncertain at present.

Figure 4 shows response rates by the number of contact attempts during the interview process: 1, 2, 3 to 4, and 5+ contact attempts.

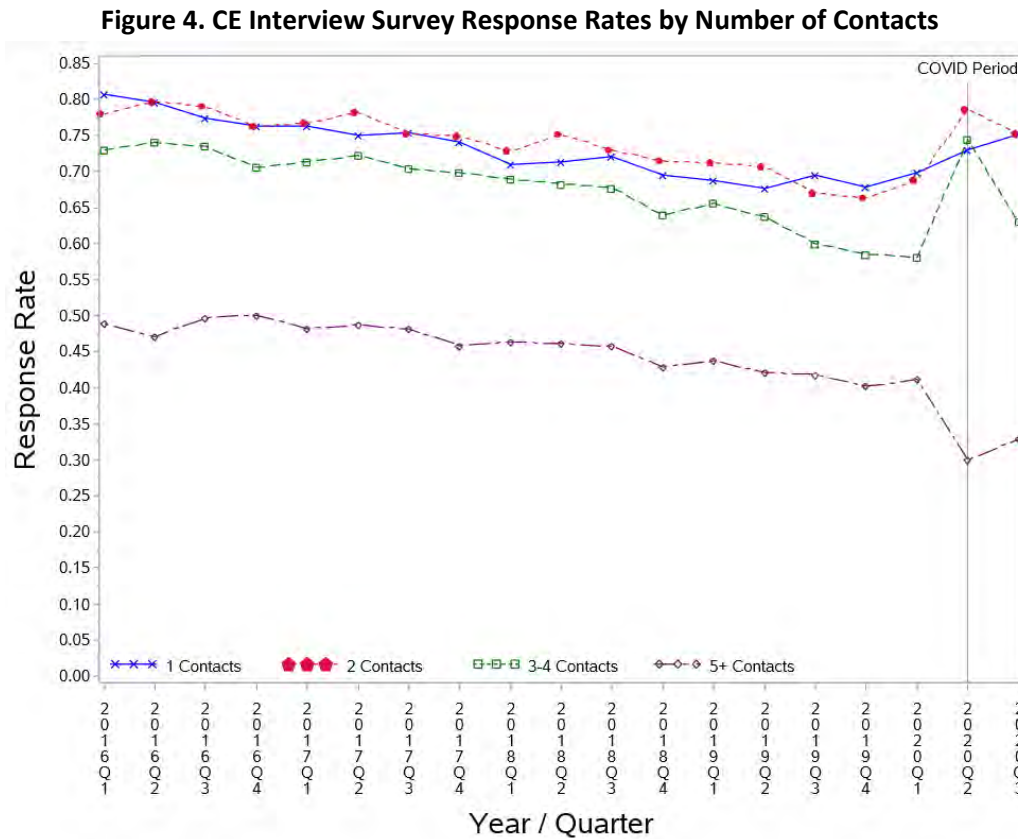**Figure 4. CE Interview Survey Response Rates by Number of Contacts**



Figure 4 suggests that the response rates are negatively correlated with the number of contact attempts. CUs with 1 or 2 contact attempts have the highest response rates, and their response rates are very similar or identical to each other in all periods examined. CUs with 3-4 contact attempts have somewhat lower response rates. However, CUs with 5+ contact attempts have much lower response rates. The strong separation of the 5+ group from the other groups suggests that the number of contacts has been, and still is, a good predictor of response rates.

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

**Figure 5: Response Rates by Census Regions**
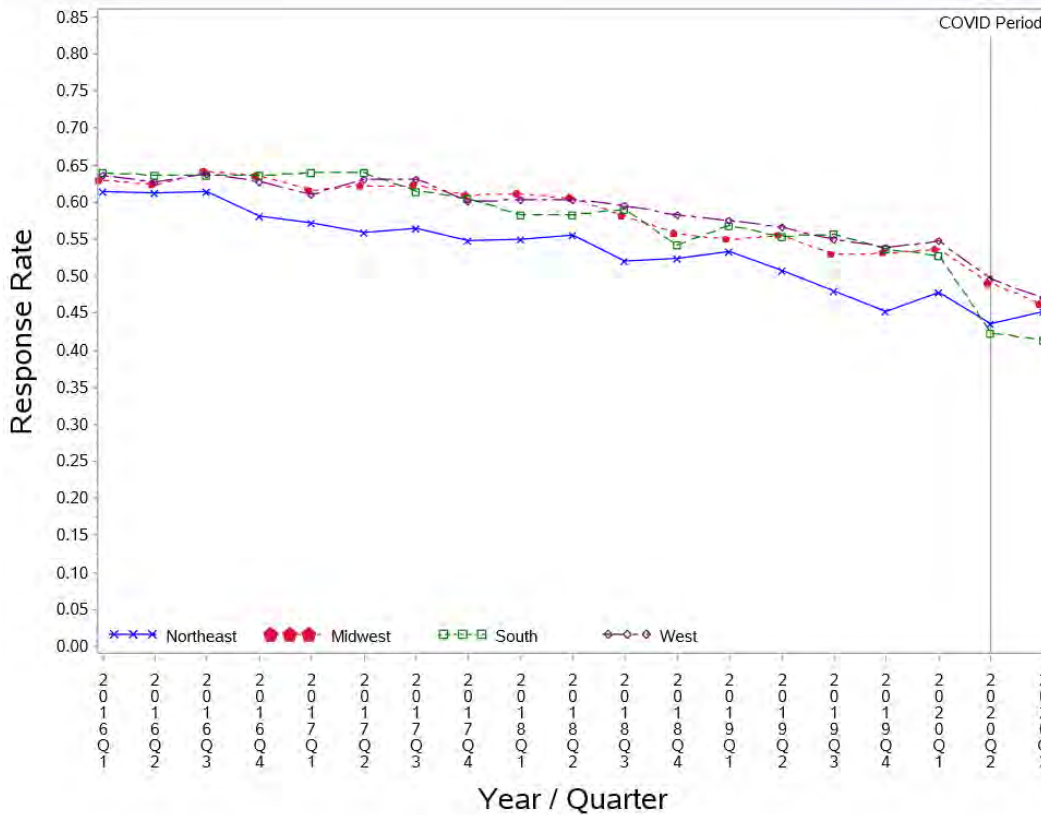


We see in Figure 5 that the Northeast has historically had the lowest response rates. However, during the COVID-19 period, the response rates for the South were lower than those for the Northeast and other regions.

**Conclusion for Analysis 1**. The response rates for the different variables used in the noninterview adjustment sometimes exhibited different patterns for before and during the COVID-19 period. This exploratory evidence suggests that the expected relationships between the response and the variables used in the noninterview adjustment has 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 an Alternative Nonresponse 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 nonresponse adjustment tailored specifically to the COVID-19 period. We suggest that the estimates from the tailored nonresponse adjustment have less bias than the current production nonresponse adjustment, and that the differences in estimates produced by the two

9

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 nonresponse adjustment for the COVID-19 period, we:

> **Examined several potential variables to see whether the nonresponse adjustment process can be improved.** We want the best possible alternative nonresponse adjustment for our comparisons. To do this, we considered all the variables used by the current nonresponse 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. Table 1 lists all the variables we considered.[3]

> **Applied a different method for calculating the nonresponse 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 was used as the nonresponse 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 1 lists the variables that we considered for the alternative nonresponse adjustments. The first and second columns of Table 1, respectively, are the variable names and descriptions of the variables considered. The third column indicates whether the variable is used in the current nonresponse adjustment, and the fourth, fifth, and sixth columns indicate whether the variable was used in our alternative nonresponse adjustments.

---

[3] 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 U.S. 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.

**Table 1: Variables Considered for Alternative Nonresponse Adjustment**

| Variable | Source | Description | Current Method | A1 | A2 | A3 |
|---|---|---|---|---|---|---|
| REGION | CE Sampling Frame | Census Region (4 categories) | X | | | |
| DIVISION | CE Sampling Frame | Census Division (9 categories) | | | X | X |
| WTIRSINC | Internal Revenue Service | IRS Income by zipcode, (Top 10%, Middle 80%, bottom 10%) | X | | | |
| WTNUMCNT | CE Administrative Data | Number of Contacts (1, 2, 3-4, 5+) | X | X | X | X |
| WTMEMQ | CE Interviewing Process | Number of persons in CU (1, 2, 3-4, 5+) | X | X | X | X |
| WTTENURE | CE Interviewing Process | Tenure (owner/renter) | | | X | |
| 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] | | | X | |
| Q1_65PLUS | Census Planning Database | Quartiles of percent of population within tract aged 65 years or older [Pop_65plus_ACS_09_13] | | | X | X |
| 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] | | | X | |
| 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 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 nonresponse 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

---

[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.

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 Table 1 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 Interview 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 aged 65 years old or older.

**Figure 6. CE Interview Survey Response Rates by Quartiles of the Census Planning Database Estimate of the Percentage of Tract Aged 65+**
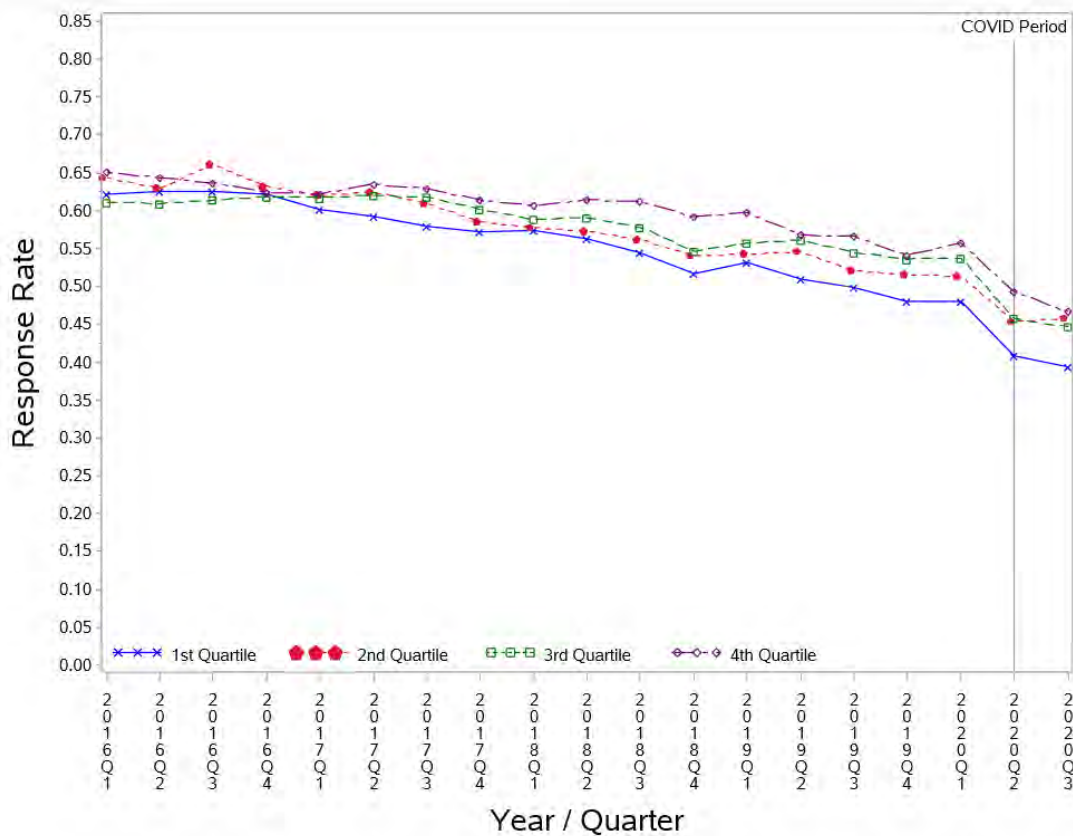
Figure 6 shows that response rates increase as the percentage of the population aged 65 years old or older increases. We see that this relationship is also present in the COVID-19 period. This suggests that the quartiles by age 65 years old or older may be a good variable in modeling response rates both before and after the COVID-19 period.

**Selection of Variables for the Alternative Nonresponse Adjustments**

The last three columns of Table 1 identify the three alternative models of the alternative nonresponse adjustments. We included these models for different reasons which we now discuss.

**(A1): WTNUMCNT WTMEMQ** – These two variables were the strongest predictors of nonresponse in all the logistic regression models that we examined; they are included in the current nonresponse adjustment; and they compose a simple model. We say that they were the strongest predictors in the logistics regression models because they consistently had the smallest p-values in all the models that we considered.
**(A2): WTNUMCNT WTMEMQ WTTENURE DIVISION INT SR_NSR Q1_65PLUS Q1_Not_HS** – This is our best model for 2020Q2 because all these variables were significant in the model for 2020Q2.
**(A3): WTNUMCNT WTMEMQ WTTENURE DIVISION INT SR_NSR**– This is our best model for 2020Q3 because all these variables were significant in the model for 2020Q3.

Research by Sabelhaus et al. (2015) showed that income was a predictor of response rates for CE. Surprisingly, none of the income variables that we considered, neither IRS income (WTIRSINC) nor quartiles of income (Q1_MED_INC), were significant after considering all the other variables in the model. We are not disputing the conclusion of Sabelhaus et al. (2015); we found that income was less important than the other variables in models (A2) and (A3).

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 nonresponse adjustment in the corresponding alternative weight. Figure 7 presents the distributions of the four nonresponse adjustment factors as applied to the completed interviews of 2020Q2. The distribution on the top-left side of Figure 7 is the distribution of the current nonresponse adjustment factors; the distribution on the top-right side is the distribution of the alternative nonresponse adjustment factors from model (A1); the distribution on the bottom-left side is the distribution of the alternative nonresponse adjustment factors from model (A2); the distribution on the bottom-right side is the distribution of the alternative nonresponse adjustment factors from model (A3).

**Figure 7. Distributions of the 2020Q2 Alternative Nonresponse Adjustment Factors**
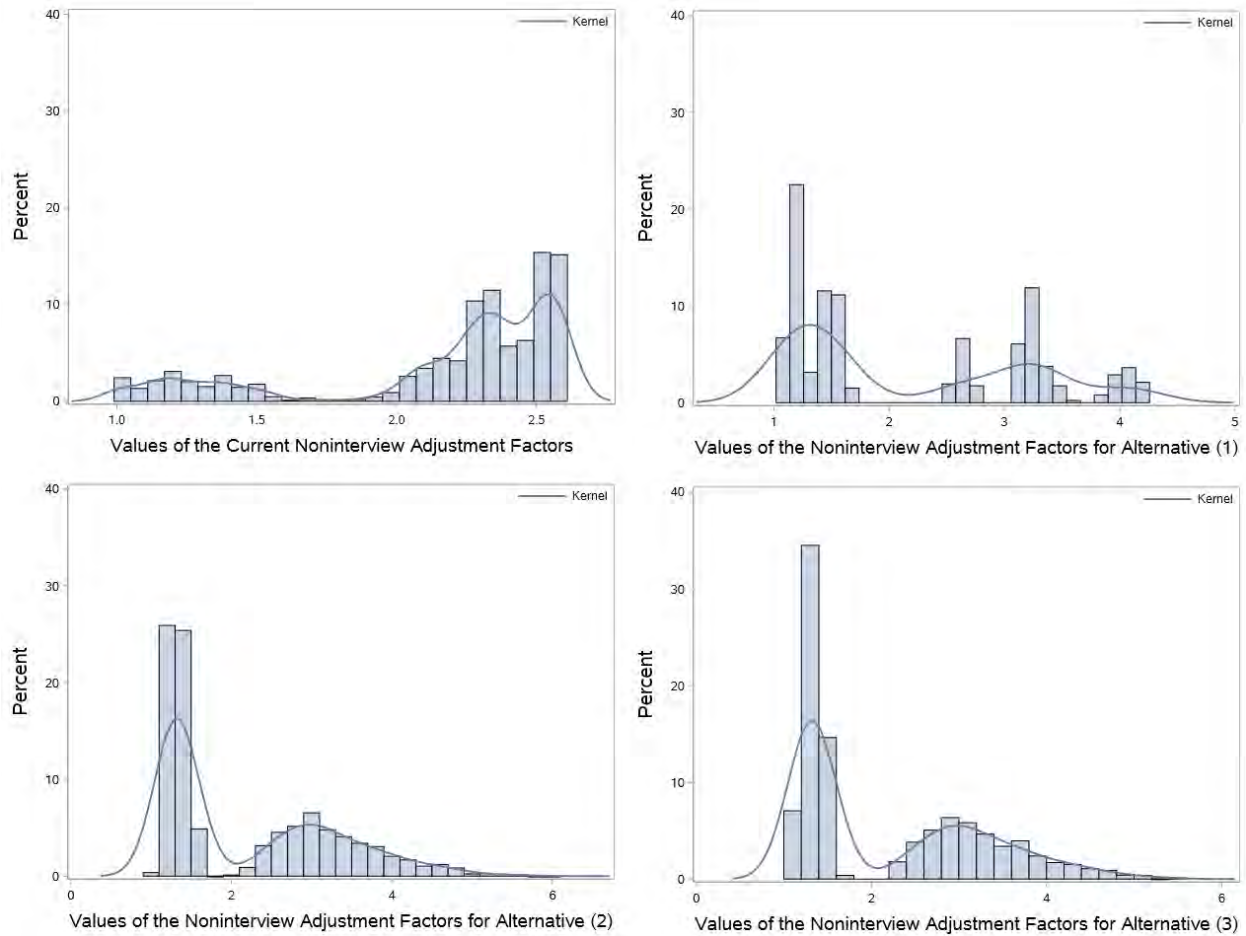


We provide the following two comments to assist the reading in interpreting Figure 7.

1) The current method caps the maximum value of the nonresponse adjustment at 2.6, which accounts for the large number of large values of the nonresponse adjustment factor. We did not cap the nonresponse adjustments for the alternative methods because the reason for capping the maximum is to reduce variances but has the effect of increasing bias. We allowed the alternative factor to be uncapped because we wanted to remove as much bias as possible.

2) The distributions of the alternative nonresponse adjustment factors are distinctly bimodal, meaning that it has two maxima or "bumps." The distribution of the current method has two maxima, but its maxima are not nearly as large as those in the alternative methods. We considered this further in Figure 8.

14

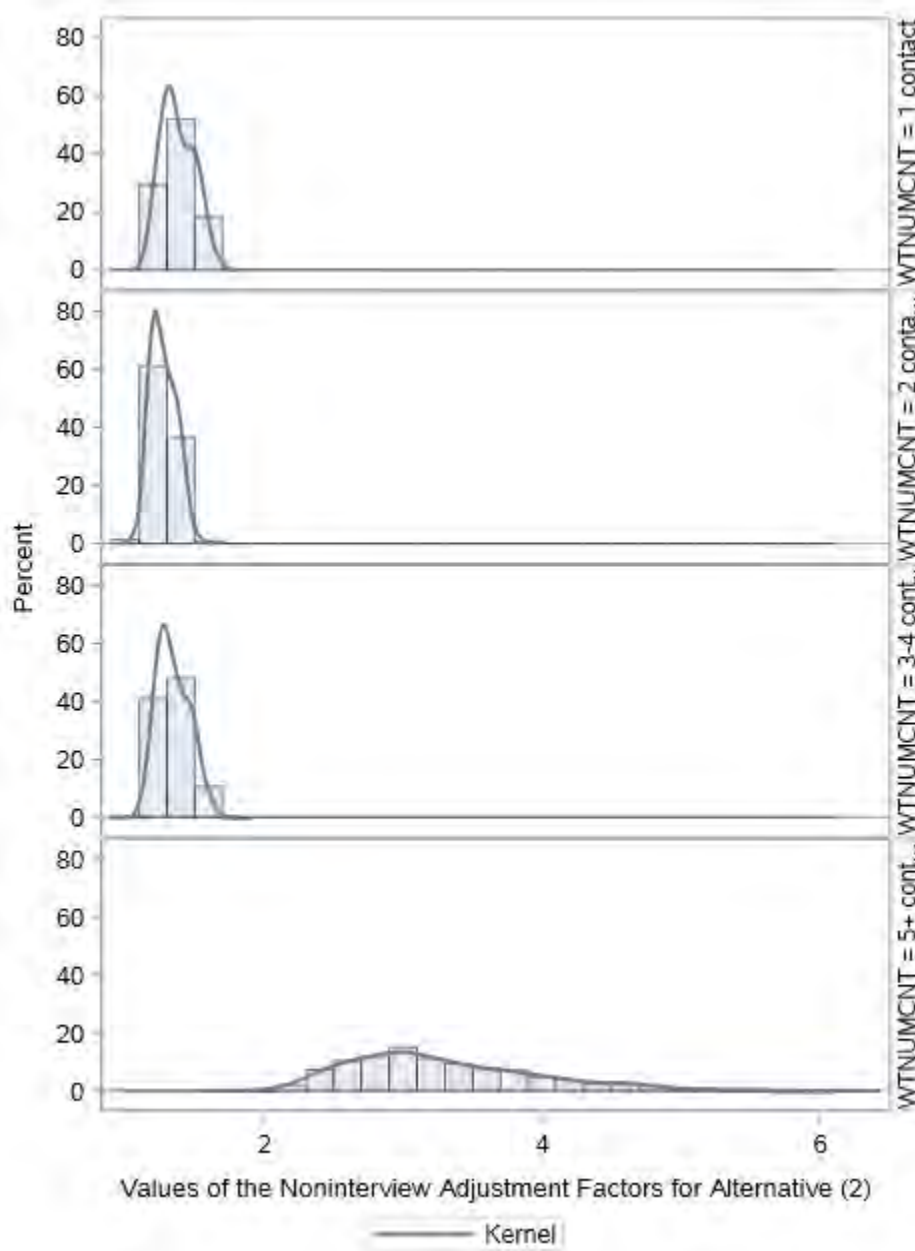**Figure 8: Distribution of the 2020Q2 Alternative Nonresponse Adjustment Factors**



Figure 8 includes four separate graphs of the distribution of 2020Q2 nonresponse adjustment factors for model (A2) by the four values of WTNUMCNT. We see that the values of the nonresponse adjustments are much different for sample CUs with 5+ contact attempts than for 1, 2, or 3-4 contact attempts. The strong relationship between the number of contact attempts and response greatly shapes the overall distribution of nonresponse adjustment factors.

The bimodal distribution of factors is a good thing because it shows the CUs were successfully stratified into sub-samples having different probabilities of responding to the survey.

**Calibration of Weights**. After the nonresponse adjustment, the last step of the weighting process is the calibration of the weights to a set of control totals. Calibration has several purposes: make the estimates agree with the control totals, reduce variances, and improve coverage, if there is under coverage. 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).

**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 using the current nonresponse adjustment and the three alternative nonresponse adjustments are a measure of nonresponse bias.

Table 2 presents the 2020Q2 national estimates of mean total expenditures for several large categories of expenditures.

**Table 2: 2020Q2 Estimates from Alternative Nonresponse Adjustments**

| | Current Method | Alternative Method (A1) | | Alternative Method (A2) | | Alternative Method (A3) | |
|---|---|---|---|---|---|---|---|
| Mean of … | Estimate | Estimate | Percent Difference | Estimate | Percent Difference | Estimate | Percent Difference |
| Total Expenditures | 56,314 (848) | 56,817 (845) | -0.9 | 56,519 (781) | -0.4 | 56,668 (832) | -0.4 |
| Housing Expenditures | 19,563 (268) | 19,777 (294) | -1.1 | 19,672 (273) | -0.6 | 19,705 (281) | -0.6 |
| Transportation Expenditures | 9,399 (490) | 9,518 (532) | -1.3 | 9,420 (498) | -0.2 | 9,454 (517) | -0.2 |
| Food Expenditures | 8,536 (115) | 8,561 (108) | -0.3 | 8,549 (106) | -0.2 | 8,557 (107) | -0.2 |
| Personal Insurance Expenditures | 6,583 (139) | 6,611 (130) | -0.4 | 6,586 (121) | -0.0 | 6,615 (128) | -0.0 |
| Health Expenditures | 4,867 (106) | 4,872 (113) | -0.1 | 4,850 (110) | 0.3 | 4,854 (109) | 0.3 |

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

The second column of Table 2 shows the estimates using the current nonresponse adjustment with their standard errors in parentheses. The remaining columns show the results of the three alternative nonresponse adjustments (A1), (A2), and (A3). For each alternative, we provide the estimate using the alternative nonresponse adjustment and the percent difference between the current and the alternative.

The magnitude of the percent differences in Table 2 are all less than or equal to 1.3 percent, which suggests there is no evidence of nonresponse bias.

Table 3 presents the 2020Q3 national estimates of mean total expenditures for several large categories of expenditures.

**Table 3: 2020Q3 Estimates from Alternative Nonresponse Adjustments**

| Mean of … | Current Method | Alternative Method (A1) | | Alternative Method (A2) | | Alternative Method (A3) | |
|---|---|---|---|---|---|---|---|
| | Estimate | Estimate | Percent Difference | Estimate | Percent Difference | Estimate | Percent Difference |
| Total Expenditures | 57,455 (636) | 58,416 (688) | -1.7 | 58,208 (712) | -1.3 | 58,428 (701) | -1.7 |
| Housing Expenditures | 20,294 (250) | 20,699 (282) | -2.0 | 20,608 (297) | -1.5 | 20,700 (282) | -2.0 |
| Transportation Expenditures | 9,503 (332) | 9,630 (349) | -1.3 | 9,618 (357) | -1.2 | 9,614 (357) | -1.2 |
| Food Expenditures | 8,528 (98) | 8,624 (101) | -1.1 | 8,604 (104) | -0.9 | 8,624 (101) | -1.1 |
| Personal Insurance Expenditures | 6,339 (132) | 6,431 (141) | -1.4 | 6,405 (138) | -1.0 | 6,444 (140) | -1.7 |
| Health Expenditures | 4,949 (116) | 4,967 (110) | -0.4 | 4,952 (113) | -0.0 | 4,964 (111) | -0.3 |

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

With Table 3, we see that magnitude of the 2020Q3 percent differences are all less than or equal to 2.0 percent, which is larger than those in 2020Q2. Although, the magnitude of the percent differences increased, we still conclude that there is no evidence of nonresponse bias.

**Conclusion for Analysis 2**. Estimates for 2020Q2 and 2020Q3 show no evidence of nonresponse bias.

**Analysis 3: Comparison of Estimates during the COVID-19 period with past CE Estimates and ACS Estimates**

The goal of this analysis is to compare CE estimates with ACS estimates during the COVID-19 pandemic. We use ACS as a benchmark because it is a well-known survey whose estimates are considered accurate. It also has a large sample size, which gives its estimates small standard errors; 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.

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 or the relative measure of the size of different subgroups. Values close to 1.0 for the relativity measure mean 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 because we know the totals from different surveys usually never 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 urban households in the U.S. may differ, however, both surveys should produce similar estimates of the proportion of households that are urban.

We use the fictitious example of Table 4 to further illustrate our measure of relativity.

**Table 4 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 | |

In Table 4, we say that CE's estimate of the proportion of the population who are owners and renters is 0.60 and 0.40, respectively. And say that ACS's estimate of the proportion of the population who are owners and renters is 0.65 and 0.35, respectively. Then 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 5 provides the CE estimates and the ACS source tables that we used in our comparisons.

**Table 5: 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)[5] (B19001) |

The value in parentheses in the second column of Table 5 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. Because 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 9 shows the relativity measures for age.
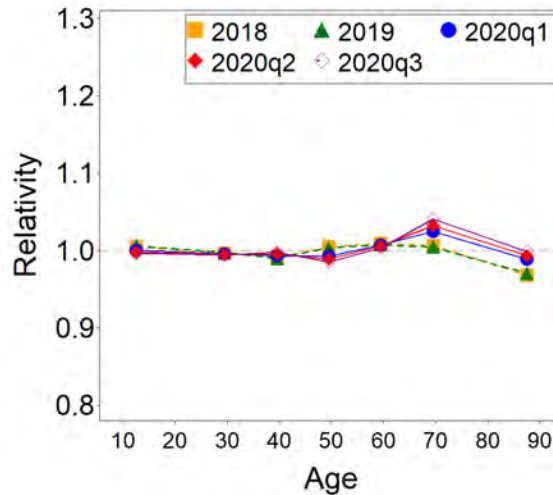


**Figure 9: Relativity Measures by Age**

The relativity measures in Figure 9 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

---

[5] 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.

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 10 shows the relativity measures for tenure.

**Figure 10: Relativity Measures by Housing Tenure**
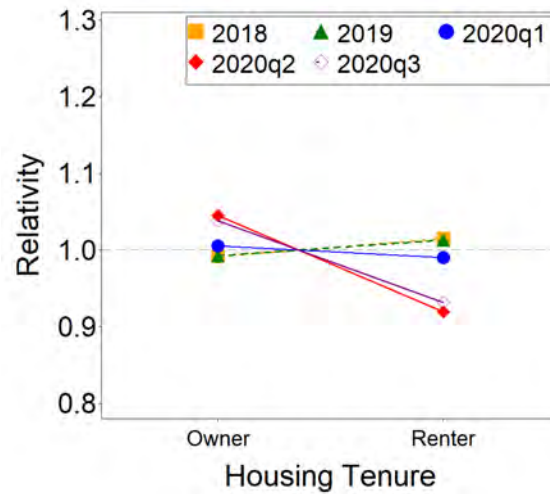


The relativity measures in Figure 10 for housing tenure are very close to 1.0 for 2018 and 2019 which means the CE and ACS estimates are in close agreement. However, there are differences between 2020Q2 and 2020Q3. Beginning with the COVID period in 2020Q2, the CPS had lower than normal response rates and higher than normal rates of homeownership than in previous quarters. The strong correlation between the change in the response rate and the change in the homeownership rate suggests that some or all of the differences are due to the CPS response rates. 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 the first quarter of 2020 and especially the second quarter of 2020. The suspension of in-person interviews affected the entirety of the data collection period for the second quarter during 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 then it decreased 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 their 2020Q2 population counts and noticed that there was an unexpected increase in the number of owner households by more than 3 million (nearly 4 percent) from the prior quarter, 2020Q1. 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 11 shows the relativity measures for size of the CU.

**Figure 11: Relativity Measures by Size of Consumer Unit**
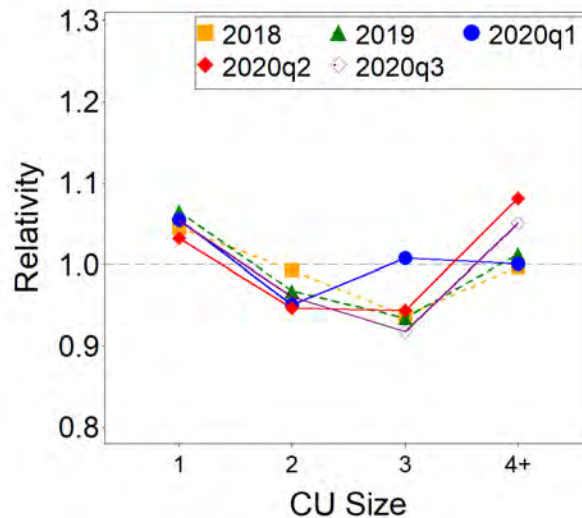


The relativity measures in Figure 11 are not as close to 1.0 as the relativity measures for age (Figure 9). This is most likely due to calibration: the CU Size variable is not used in calibration but is used in nonresponse adjustment. During the COVID-19 period, it appears that the relativity measure for CUs with 1 person is much closer to 1.00 than the recent prior periods while the 4+ persons CUs has drifted a bit further from 1.00 in the COVID-19 period.

Figure 12 shows the relativity measures for the level of education.

**Figure 12: Relativity Measures by Level of Education**
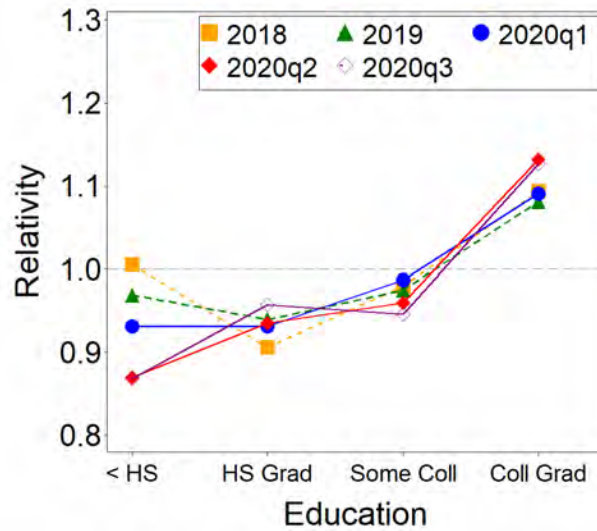


The relativity measures by level of education in Figure 13**Error! Reference source not found.** show that CE and ACS are not always in agreement, but we do see that the COVID-19 period has drifted away from unity for less than high school and for college graduates when compared to 2018 and 2019.

Figure 13 shows the relativity measures by the income of the CU.
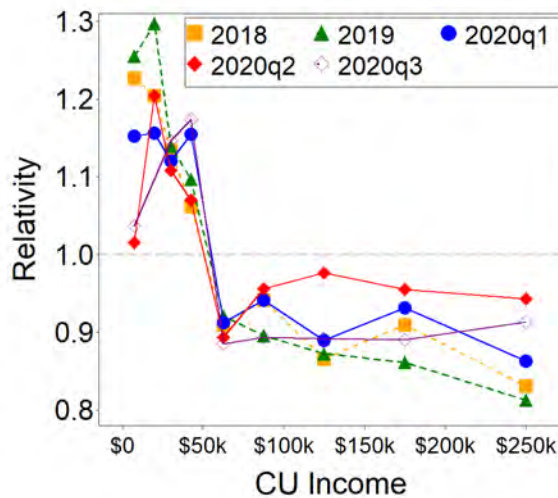
**Figure 13: Relativity Measures by Income**



Figure 13 shows that the relativity measures for the COVID-19 period did no worse than in 2018 and 2019. In fact, for estimates of high-income households, and to lesser degree lower-income households, the relativity measures for the COVID-19 period were closer to 1.0 than 2018 and 2019. We also

speculate that the larger values of the relativity measures are possibly due to differences in how CE and ACS define income.

**Conclusion of Analysis 3**. In most cases, we found that the CE estimates during the COVID-19 period were not greatly different than the ACS estimates. The one exception was estimates of housing tenure that was discussed above. In some cases, like size of the CU and income, the relativities for the COVID-19 period were better than 2018 and 2019.

**Overall 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 sometimes exhibited different patterns for before and during the COVID-19 period. In Analysis 2, we computed expenditure estimates with three alternative weighting procedures that were tailored to the COVID-19 situation, 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 from 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**

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. (2009). 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

Knappenberger, C., Lee, Y., Pham, S., and Armstrong, G. (2021a). "Reference Guide CE Data Quality Profile," Consumer Expenditure Surveys Program Report Series, April 23, 2021. https://www.bls.gov/cex/dqp-reference-guide-midyear-2020.pdf

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

Krieger, S., McBride, B., Nix, B., Steinberg, 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 2007-2016," *forthcoming*.

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. & 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., 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. (2020). U.S. Bureau of Labor Statistics requirement "Phase 2 Interview Production Requirement – Weighting, 2020 Release, version 1.0," dated May 11, 2020.

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 March 28, 2018.
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

**Response Rates by Selected Characteristics**

**Figure A1: Response Rates by Tenure Status**
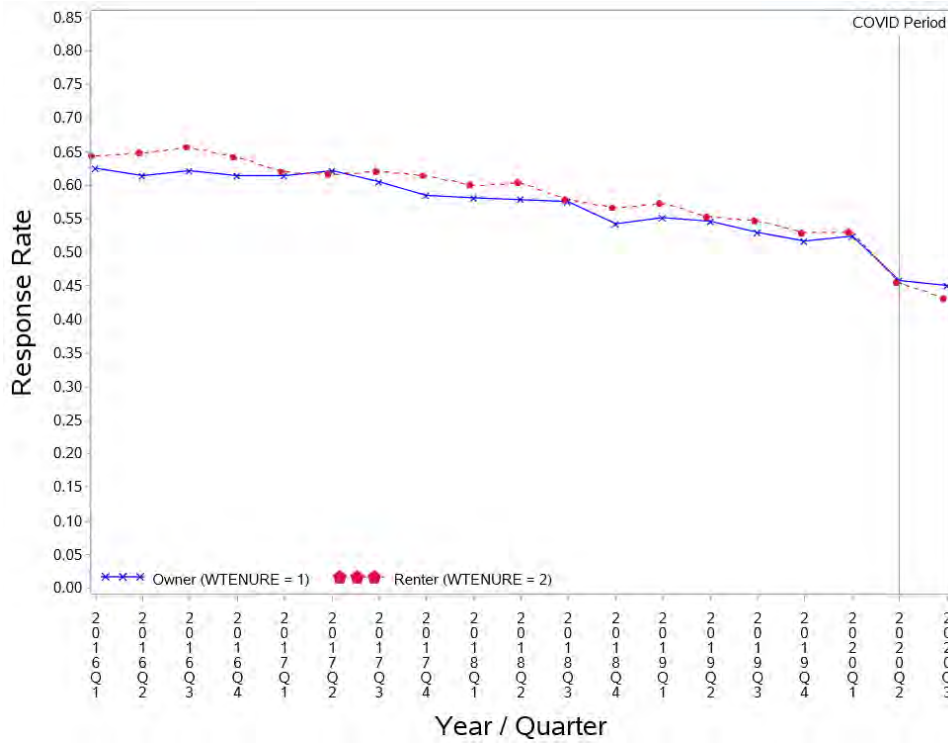


**Figure A2: Response Rates by Interview Number**
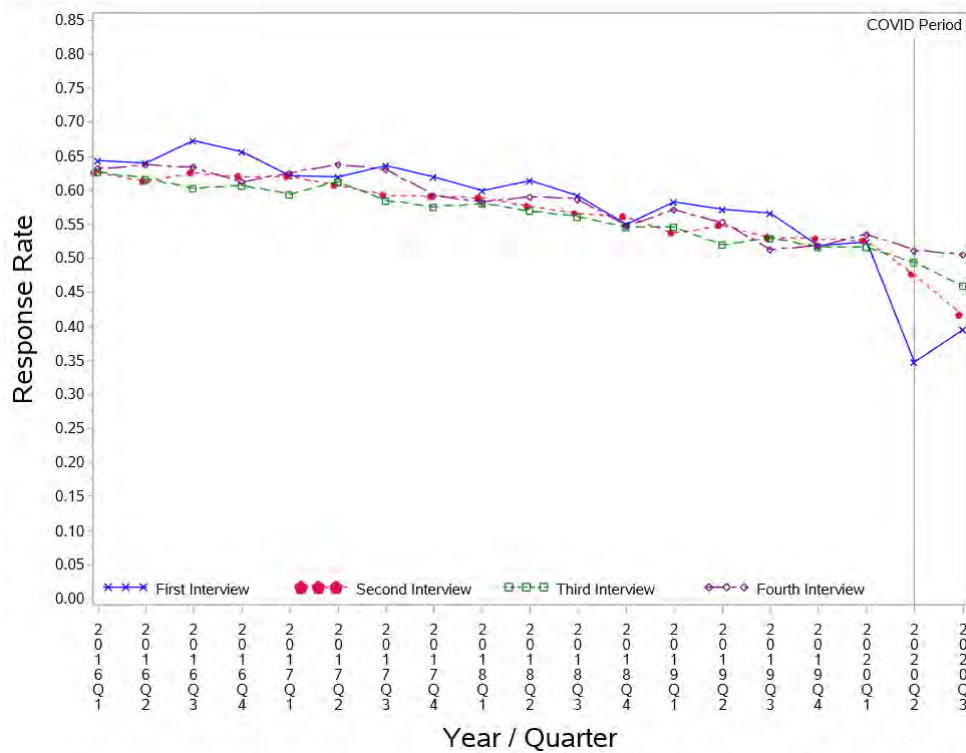
**Response Rates by Selected Characteristics**

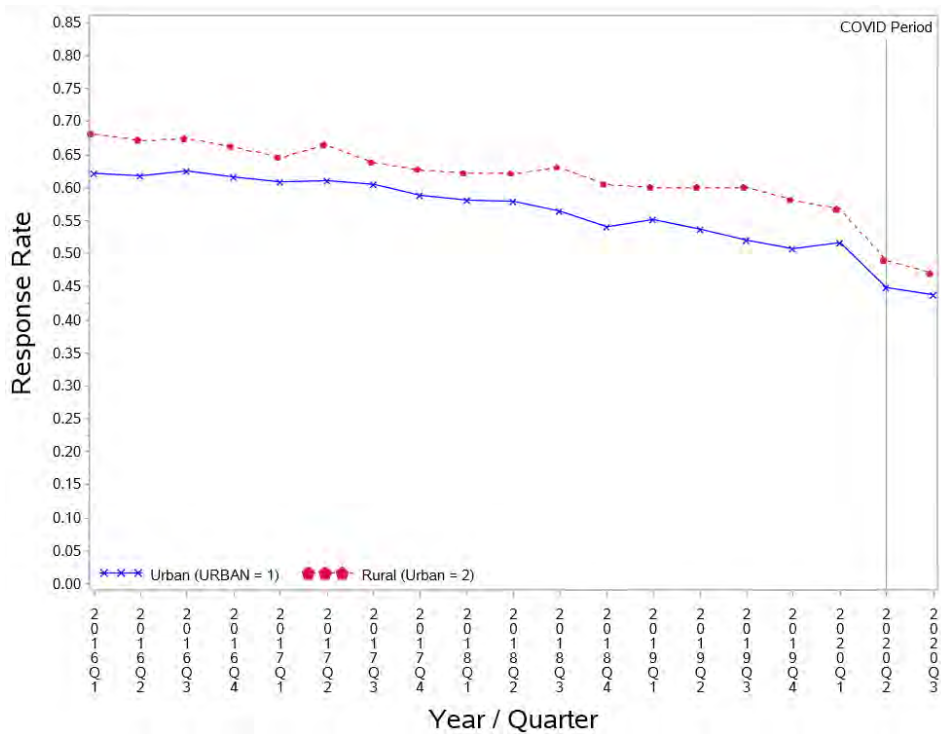**Figure A3: Response Rates by Census Division**
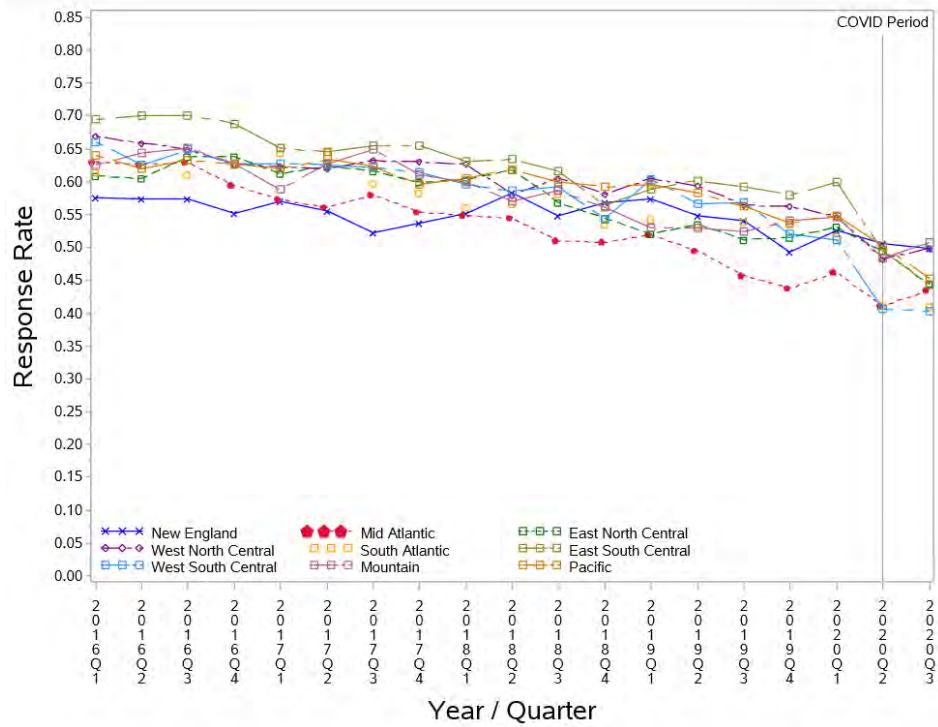


**Figure A4: Response Rates by Urban/Rural**

**Response Rates by Selected Characteristics**
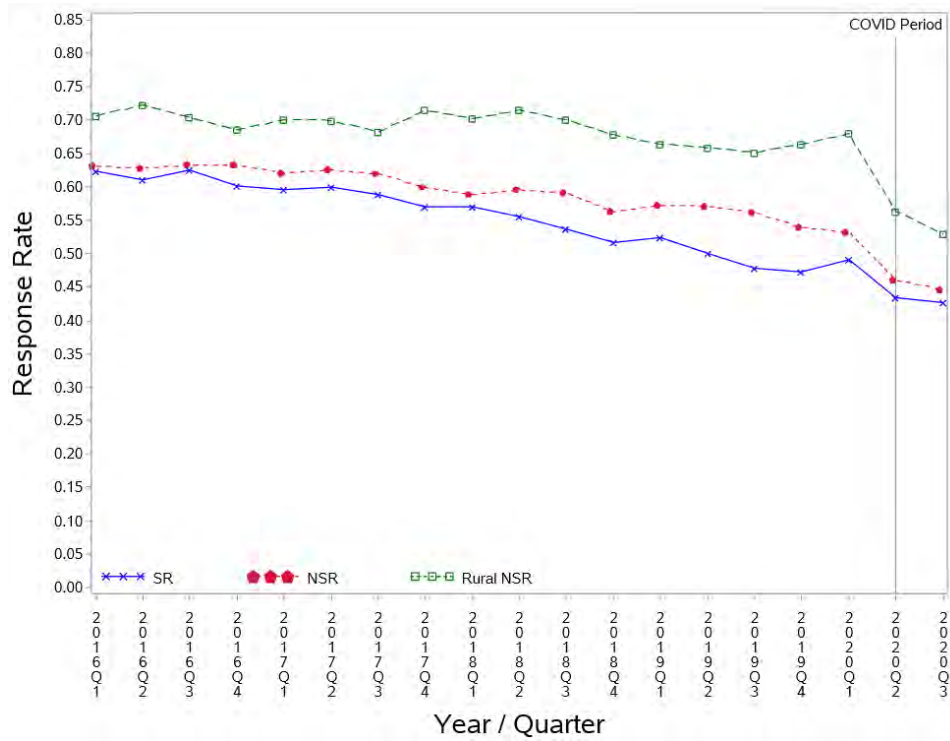
**Figure A5: Response Rates by Size of Strata**



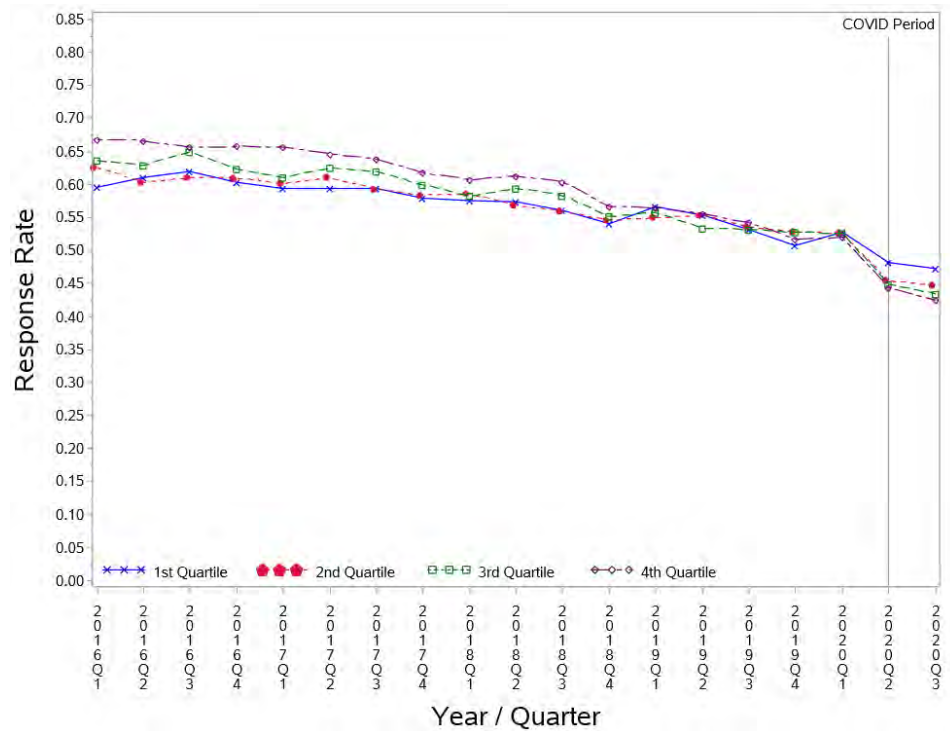**Figure A6: Response Rates by Quartiles of Percent Not High School Graduate Population**

**Response Rates by Selected Characteristics**
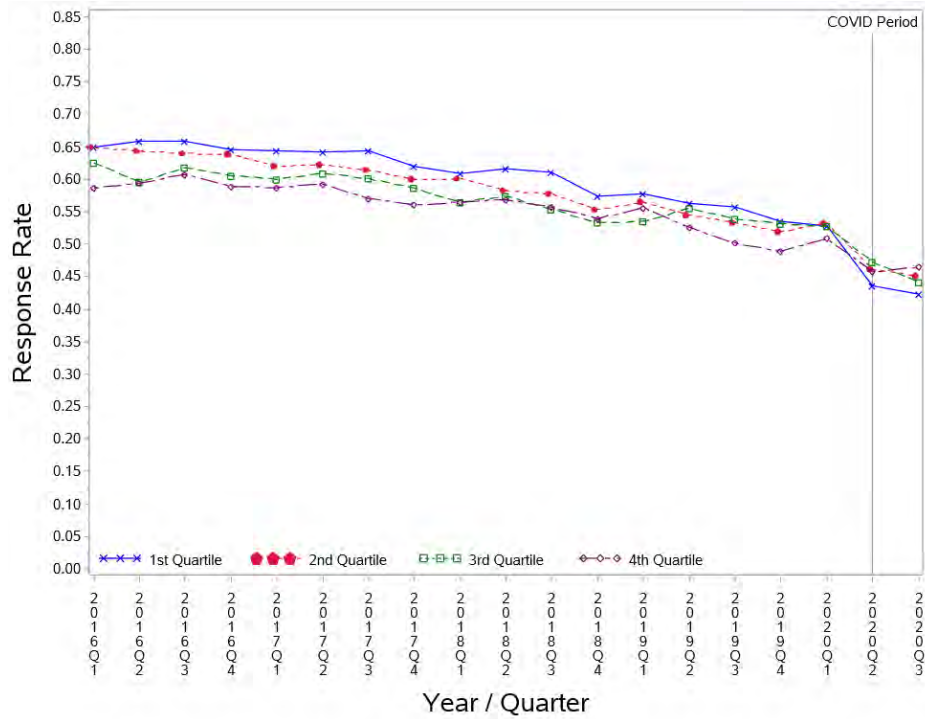
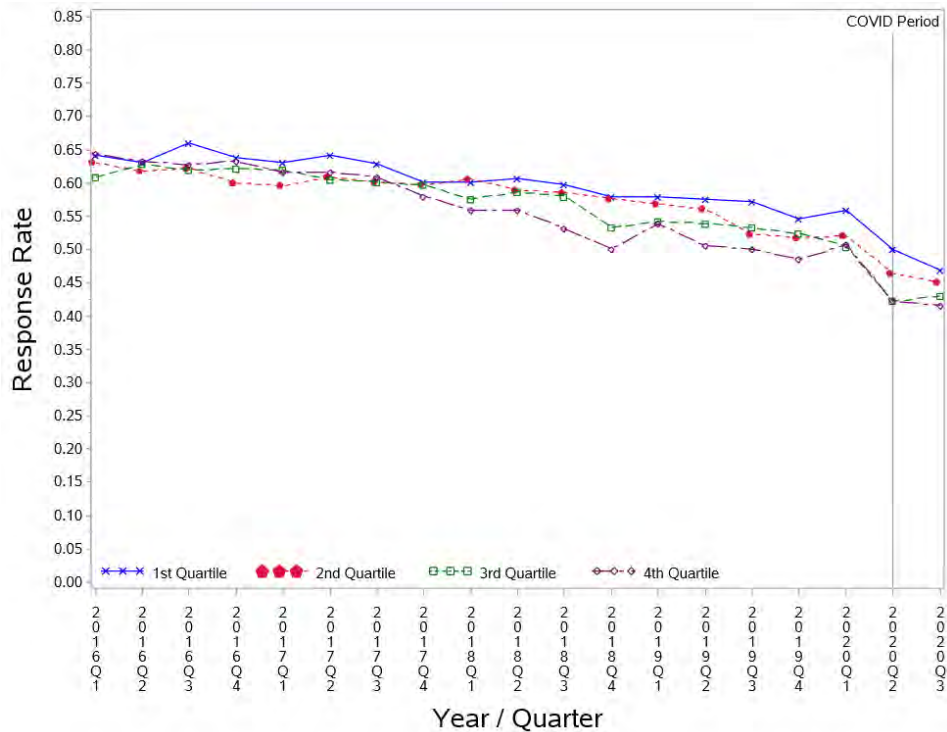**Figure A7: Response Rates by Quartiles of Median Income**



**Figure A8: Response Rates by Quartiles of Percent Black Alone Population**

**Response Rates by Selected Characteristics**

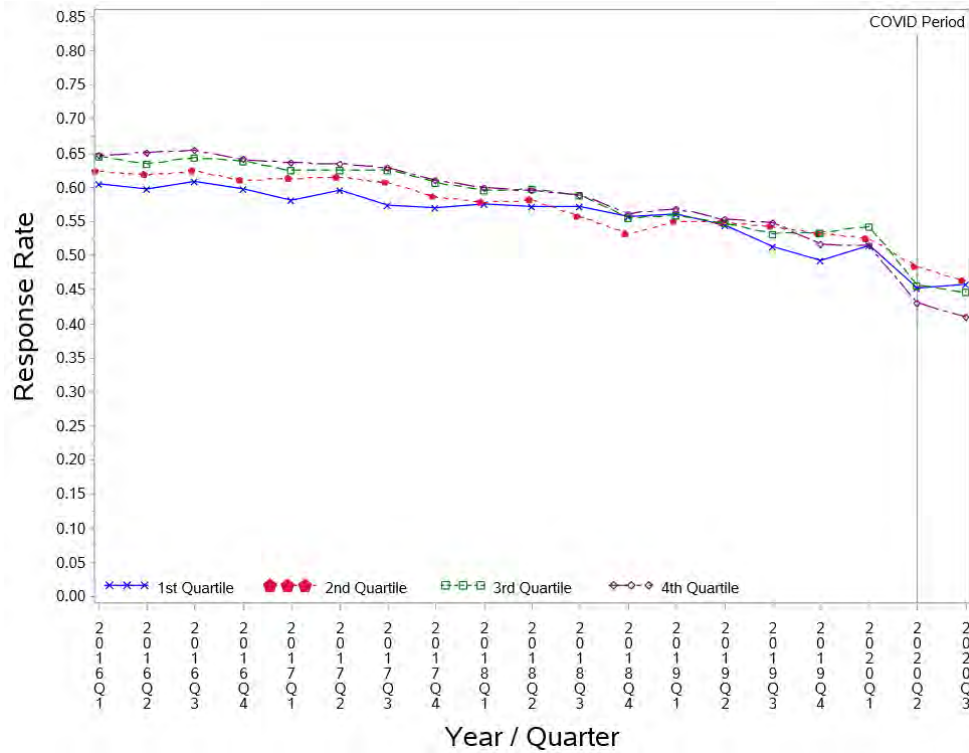**Figure A9: Response Rates by Quartiles of Percent Households under Poverty Level**



**Figure A10: Response Rates by Quartiles of Percent Hispanic Population**