



COVID-19 CE DIARY PRELIMINARY REPORT: Response and Data Quality Impacts

Brett McBride & Nikki L. Graf

Melissa Edwards & Meaghan Smith, Consultants

Consumer Expenditure Surveys Program Report Series

I. Introduction

Following the outbreak of the COVID-19 virus in the U.S. in early 2020, the Census Bureau was forced to adopt a number of measures to revise data collection for surveys, including for the Consumer Expenditure (CE) Diary Survey. Normal diary collection involved respondents handing over diaries with two weeks of their expenditures to the interviewer. Given quarantine and other pandemic measures, starting around March 19, 2020, the Census Bureau suspended personal visits, and interviewers instead began calling to recruit respondents and verbally collect expenditures (twice over the two-week period) with the aid of receipts and a diary form sent by mail or email. In a second period, starting in June 2020, interviewers started to provide URLs to English-speaking respondents who access the internet at least a few times a week to enable them to enter expenditures in an online diary. For those without internet or who preferred not to use an online diary, interviewers used phone transcription to record expenditures. This report documents how the pandemic and these revised data collection procedures affected spending behavior, diary reporting, and data quality in the first half of 2020.

II. Data and Methods

In this report, we examined data from January 2020 through June 2020, which cover months included in CE's midyear data release. The pandemic fully affected the last part of the data we examined (March 19 through June 30, 2020). The data we analyzed mostly consisted of those that had undergone some editing. To assess the effect of COVID-19 on data quality, the analysis items compared data segmented into the following periods that align with changes in Census data collection protocols:

- **Baseline** – January 1-March 18, 2020.¹ This period involved normal procedures for diary placement and interviewer pickup. With few exceptions, most diaries involved normal respondent entry procedures. This was considered the baseline dataset for comparison with subsequent COVID-19-affected time periods.
- **Period 1** – March 19-May 31, 2020. In this period, personal visits ceased, and diary collection involved interviewers calling respondents to collect expenditures respondents had been tracking.

¹ Any household with a diary placed in March with a pick-up date preceding March 19 was in the baseline period. (Six households with diaries placed in February that had pick-up dates on March 19 or later were also in the baseline period based on their diary keeping time period).

Interviewers then transcribed entries into a paper diary. Period 1 also included phone transcription from a segment of completed paper diaries that interviewers could not pick up. The Bureau of Labor Statistics also reclassified some outcomes to designate COVID-19-related nonresponse.

- Period 2 – June 1-30, 2020. In period 2, interviewers referred eligible respondents to online diaries with an alternative of phone transcription.²

A subsequent report will examine data from the second half of 2020, during which there was a phased resumption of in-person diary placement and pickup, in addition to the collection modes described above. For the data quality analysis in this report, we subset diary weeks to Consumer Units (CUs) considered complete for publication.³

The counts of CUs completing the interview in each time period are shown below (Table 1).

Table 1. Sample of CUs by time period

Month (2020)	Baseline	Period 1	Period 2	Total
January	614	-	-	614
February	594	-	-	594
March	60	399	-	459
April	-	334	-	334
May	-	341	-	341
June	-	-	318	318
Total	1,268	1,074	318	2,660

CUs having at least one week with a complete diary disposition did not always provide expenditures for both weeks of diary reporting, resulting in the diary counts shown below (Table 2).⁴

Table 2. Count of diaries having a complete disposition with entry status by time period and mode

	Diaries	With entries	No entries
Baseline	2,496	2,402	94
Period 1	2,097	2,037	60
Period 2	628	600	28
Telephone	424	401	23
Online	204	199	5

² There were six respondents receiving personal visits in period 2. However these were not categorized as a separate mode as they only involved a personal visit for diary pickup and assumedly involved diary recording processes similar to others in period 2 (e.g., online for one and using phone transcription for the others).

³ A Consumer Unit consists of families, single persons living alone or sharing a household with others but who are financially independent, or two or more unrelated persons living together who pool their income to make joint expenditure decisions. Complete diaries include CUs that were ‘temporarily absent’ during the reporting period, but do not include some diaries that have low or no reported expenditure amounts.

⁴ Diaries with no entries could occur based on survey editing (e.g., if the CU was temporarily absent during the reporting period, in which case expenditures do not reflect the location of their sampled address).

The assessments of the effects of COVID-19 on the Consumer Expenditure Diary Survey are based on the following analyses and methods:⁵

1) Demographic Analysis. We investigated whether COVID-19 and related procedural changes had an impact on the composition of the participating sample, which could have affected data quality. We examined demographic characteristics at the CU-level, including CU size, housing tenure, and the following characteristics of the reference person:⁶ age, gender, race, Hispanic origin, and educational attainment.

2) Spending Changes: Legitimate Blank Diaries. To examine the lack of expenditures among the CUs we analyzed, we calculated the proportion of complete diaries that were legitimate blank diaries. Among those interviews considered complete by processing, a legitimate blank diary was indicated by the CU reporting to the interviewer that they did not have expenditures that week or, for online diaries, by the CU marking a checkbox indicating no expenditures. Legitimate blank diaries included only those diaries where the calculated number of entries was zero (also referred to as a ‘blank diary’).

In addition to analyzing legitimate blank diaries across time periods, we analyzed differences by mode for diaries in period 2.

3) Expenditure Counts: Section and Mode. One indicator of a potential reduction in data quality would be less comprehensive reporting behavior. We examined this through statistics on the number of entries by diary week, broken out by section. To examine the impact of different modes on reporting, we broke out weekly expenditure counts by mode in period 2. For the calculation of overall diary counts we subset to diary weeks containing one or more entries, but for section-level analysis, we created zero counts to account for sections with no expenditures. Some other factors not examined in this report, such as CUs receiving the initial round of stimulus checks distributed under the Coronavirus Aid, Relief, and Economic Security Act (also known as the “CARES Act”) of 2020, may have also affected the counts of respondent entries.⁷ Additionally, we examined the extent to which interviewers prompted respondents to recall forgotten expenditures across time periods.

4) Drop-Off in Weekly Reporting Behavior. We analyzed respondent reporting behavior over the two-week period by examining the drop-off in expenditure entries over time. We measured this by calculating the percentage of two-week diary expenditure entries reported in week 1 (i.e., number of expenditure entries in week 1 over total number of expenditure entries reported over both weeks) for each CU that had at least 1 entry in each week, and then averaging that percentage across CUs. By limiting the analysis to CUs with entries in each week, we control for CUs only participating in one or the other diary weeks. Percentages greater than 50 reflected a drop-off in entries across the 2 weeks, potentially signaling a reduction in data quality. Additionally, we examined what the number of CUs with blank diary weeks suggested about reporting consistency.

⁵ For details on the methods we used, please see the appendix.

⁶ In CE, the reference person is the first member mentioned by the respondent when asked to “Start with the name of the person or one of the persons who owns or rents the home.”

⁷ “CARES” is the official acronym for the act. See “Section 1: Short Title” of H. R. 748, “An Act: To amend the Internal Revenue Code of 1986 to repeal the excise tax on high cost employer-sponsored health coverage,” 116th Congress, second session (begun 1/3/2020), <https://www.congress.gov/116/bills/hr748/BILLS-116hr748enr.pdf>.

5) Item Non-Response. We examined fields requiring a response, identifying those where the respondent did not know the answer or refused to provide one. We also analyzed the rate of item non-response in the cost and item description fields (excluding the Meals section⁸) across the time periods, and carried out significance testing by mode for diaries in period 2. (For more information on statistical methods, see the appendix).

III. Findings

1. Demographic Analysis

The share of CUs that agreed to participate in the CE Diary survey was affected by COVID-19, with increasing rates of nonresponse observed over the quarters included in the CE midyear release. More information on changes in overall nonresponse and nonresponse category (e.g., noncontact, refusal) is available in a separate document profiling changes in the quality of CE data.⁹

To explore whether COVID-19 had an impact on the composition of the responding sample because of differential nonresponse, we examined demographic characteristics (Table 5). Reference persons in periods 1 and 2 generally had higher levels of education¹⁰ and were more likely to be homeowners than those in the baseline period. For example, 65.4 percent of those in the baseline period were homeowners, compared with 71.7 percent and 74.0 percent in periods 1 and 2, respectively.¹¹ Reference persons in periods 1 and 2 also tended to be older than those in the baseline period.¹² In addition, the sample in periods 1 and 2 had somewhat lower proportions of Hispanic respondents than in the baseline period.¹³ These differences could have been related to the change from in-person to telephone recruitment, prompted by COVID-19, which required more reliance on Census resources to obtain phone numbers for sample addresses.

⁸ For this section, respondents do not enter a description beyond providing the name of the vendor.

⁹ [The Midyear CE Data Quality Profile - 2020](#)

¹⁰ For more details on the tests used to determine significance, see appendix. Test for the equality of proportions of reference persons with a bachelor's degree or higher – Baseline vs. Period 1: $Z = -2.35$, two-sided ($p = 0.02$); Baseline vs. Period 2: $Z = -1.78$, two-sided ($p = 0.08$).

¹¹ Baseline vs. Period 1: $Z = -3.26$, two-sided ($p = 0.001$). Baseline vs. Period 2: $Z = -2.89$, two-sided ($p = 0.004$).

¹² Test for the equality of proportions of reference persons ages 50 and older – Baseline vs. Period 1: $Z = -2.40$, two-sided ($p = 0.02$); Baseline vs. Period 2: $Z = -2.16$, two-sided ($p = 0.03$).

¹³ Baseline vs. Period 1: $Z = 2.74$, two-sided ($p = 0.006$). Baseline vs. Period 2: $Z = 1.81$, two-sided ($p = 0.07$).

Table 5. Sample characteristics by time period and mode (excludes CUs with no expenditure entries)

	Baseline	Period 1	Period 2	In Period 2:	
				Online	Telephone
Number of Consumer Units	1,237	1,065	315	102	213
Reference person characteristics:					
Race	%	%	%	%	%
White	80.6	81.7	85.1	81.4	86.9
Black	9.6	9.0	7.3	6.9	7.5
Other (incl. Asian, multi, other)	9.8	9.3	7.6	11.8	5.6
Hispanic origin					
Hispanic or Latino	13.7	10.0	9.8	5.9	11.7
Not Hispanic or Latino	86.3	90.0	90.2	94.1	88.3
Gender					
Female	51.2	54.2	57.1	56.9	57.3
Male	48.7	45.8	42.9	43.1	42.7
Age					
18-24	4.4	2.3	1.3	2.0	0.9
25-34	15.4	12.4	11.1	13.7	9.9
35-49	25.1	25.3	25.7	35.3	21.1
50-64	27.8	28.5	32.7	27.5	35.2
65 and older	27.3	31.5	29.2	21.6	32.9
Education					
Less than high school	7.7	5.6	7.9	1.0	11.3
High school graduate	19.6	18.7	17.1	13.7	18.8
Some college or associate's deg.	32.2	30.3	28.9	26.5	30.0
Bachelor's degree or higher	40.5	45.4	46.0	58.8	39.9
CU characteristics:					
CU size					
One person	29.6	32.5	28.3	21.6	31.5
2-3 persons	48.2	47.3	51.4	51.0	51.6
4+ persons	22.2	20.2	20.3	27.5	16.9
Housing tenure*					
Renter	33.5	28.3	25.7	21.6	27.7
Owner	65.4	71.7	74.0	78.4	71.8

*For housing tenure, college student housing CUs (13 CUs in baseline and 1 CU in period 2) are not shown.

Starting in period 2, interviewers gave eligible respondents the option of using an online diary, rather than telephone transcription, to record their expenditures. Looking at demographic differences by self-selected mode within period 2, reference persons in CUs completing online diaries were far more likely to be college graduates than those in CUs with diaries requiring telephone transcription (58.8 percent vs. 39.9 percent).¹⁴ Reference persons in the online diary group were also generally younger – 51.0 percent were under age 50, compared with 31.9 percent in the telephone transcription group.¹⁵ In addition, CUs with

¹⁴ Z = 3.15, two-sided (p = 0.002).

¹⁵ Z = 3.26, two-sided (p = 0.001).

online diaries tended to be larger (27.5 percent had 4 or more persons vs. 16.9 percent in the telephone transcription group).¹⁶

2. Spending Changes: Legitimate Blank Diaries

The impact of COVID-19 on CE diary reporting resulted from both changes in spending patterns as well as potential deterioration in the quality of data collected. Using post-processed data to maintain consistency with the analysis sample used throughout this report, we identified CUs that experienced a one-week period without expenditures.¹⁷ To examine this, we calculated the proportion of complete diaries that were legitimate blank diaries. It should be noted that some blank diaries are removed during CE processing and therefore are not included in the findings presented below.¹⁸

Among those interviews considered complete by processing, legitimate blank diaries are defined as those in which, for a given week, a CU reported to the interviewer that its members had no expenditures. Additionally, in online diaries, respondents could indicate that they had no expenses by checking a checkbox, which appeared on the last day of the week. For the online mode, diaries with no expenditure entries were counted as legitimate blank diaries if either (1) the CU reported to the interviewer that they did not have expenditures, or (2) the CU checked the “no expenses” checkbox for that week. Legitimate blank diaries include only those diaries where the calculated number of entries was zero.¹⁹

Table 6 shows the proportion of legitimate blank diaries across time periods and modes of data collection, both overall and by diary week. The share of legitimate blank diaries remained very low across the time periods analyzed. Overall, just 1.4 percent of complete diaries in the baseline period were legitimate blank diaries, compared with 1.1 percent in period 1 and 2.1 percent in period 2.²⁰ In both the baseline period and period 1, the proportion of legitimate blank diaries increased from week 1 to week 2.²¹ However, in period 2, this pattern by week is not evident – the share of legitimate blank diaries overall remained roughly stable between weeks 1 (1.9 percent) and 2 (2.2 percent).²²

¹⁶ $Z = 2.18$, two-sided ($p = 0.03$).

¹⁷ In contrast to a reduction in spending (i.e., actual dollars spent), the pandemic could have led to respondents buying larger amounts with less frequency, an outcome that would also be consistent with a given one-week period not having any expenditures recorded.

¹⁸ See the “[Reference Guide CE Data Quality Profile](#)” under the section “Final disposition rates of eligible units” discussing nonresponse reclassification for details.

¹⁹ Among complete diaries, there were 21 diaries in the baseline period, 21 diaries in period 1, and 4 diaries in period 2 that had entries, even though the interviewer reported that there were no entries and the CU reported that they did not have expenditures. Since they had entries, these diaries were not included in the counts of legitimate blank diaries in Table 6.

²⁰ These differences are not statistically significant at the .05 level. Baseline vs. Period 1: $Z = 0.66$, two-sided ($p = 0.51$). Baseline vs. Period 2: $Z = -1.31$, two-sided ($p = 0.19$). Period 1 vs. Period 2: $Z = -1.76$, two-sided ($p = 0.08$).

²¹ Week 1 vs. Week 2 in Baseline: $Z = -2.74$, two-sided ($p = 0.006$). Week 1 vs. Week 2 in Period 1: $Z = -3.71$, two-sided ($p = 0.0002$).

²² Week 1 vs. Week 2 in Period 2: $Z = -0.27$, two-sided ($p = 0.79$).

Table 6. Percent of legitimate blank diaries by time period and mode

	N (diaries)	Total	Week 1	Week 2
Baseline	2,496	1.4%	0.7%	2.0%
Period 1	2,097	1.1%	0.3%	2.0%
Period 2	628	2.1%	1.9%	2.2%
Telephone	424	2.6%	2.8%	2.3%
Online	204	1.0%	0.0%	2.0%

Within period 2, we also examined differences by mode. The share of legitimate blank diaries was 2.6 percent for diaries involving telephone transcription and 1.0 percent for online diaries.^{23 24} These data should be interpreted with caution due to the relatively small number of online diaries available for analysis at this stage, which, when combined with the low incidence of legitimate blank diaries, makes it difficult to draw conclusions. Future analysis with additional months of data involving online diaries will further explore mode comparisons, as well as patterns across weeks and what these may signal for data quality.

3. Expenditure Counts: Section and Mode

With the assumption that a higher number of diary entries, beyond signifying more actual expenditures, may indicate better data quality, we examined diary entries both at the week-level and by section, for non-blank diaries. Table 7 displays the change in entries over time and by mode of data collection.

Table 7. Count of weekly diary entries by time period and mode

	N (diaries)	Mean	Median	Min	Max
Baseline	2,402	33.1	28	1	194
Period 1	2,037	26.2	18	1	206
Period 2	600	21.1	15	1	137
Telephone	401	19.5	14	1	137
Online	199	24.3	17	1	111

We can see that, whether due to actual pandemic-induced changes in spending patterns or deterioration in reporting behavior, CE diary CUs reported a significantly lower average number of expenditures in each subsequent time period, with a 21 percent decline in period 1 (33.1 to 26.2) followed by another 19 percent decline in period 2 (26.2 to 21.1).²⁵ Although we do not control for demographic changes, this series of declines came despite an increase over time in homeowner CUs, which usually have higher expenditures than renter CUs. Within period 2, respondents recording expenditures using online diaries provided a significantly higher number of entries than those providing them through phone

²³ This difference was not statistically significant: $Z = -1.33$, two-sided ($p = 0.18$).

²⁴ The [Proof of Concept Test report \(2018\)](#) appeared to find a similar pattern by mode – the share of legitimate blank diaries was 2.6 percent among paper diaries and 0.3 percent among online diaries (there are several methodological differences between the Proof of Concept Test and the production data analyzed in this report, however). Future analysis of legitimate blank diaries may provide additional insight into these mode comparisons.

²⁵ For more details on the tests used to determine significance, see appendix. Wilcoxon Rank-Sum two sample test; Baseline to Period 1, $Z = -12.40$, two-sided ($p < 0.0001$). Period 1 to Period 2, $Z = -3.91$, two-sided ($p < 0.0001$).

transcription.²⁶ Reduced reporting in period 2 among telephone respondents could indicate the presence of underreported expenditures due to recall error among those respondents, compared to the more contemporaneous reporting made possible via online diaries. However, it is likely that the difference in size between CUs placed with online diaries (larger) and CUs involving telephone transcription (smaller) explains this gap between modes. This mode difference will be examined further as part of a future report, to see whether controlling for factors other than mode (e.g., size of CU) explains the difference. To further investigate changes in expenditure counts, we examined data by Diary section across the time periods (Table 8).²⁷

Table 8. Count of weekly Diary section entries by time period and mode

	N (diaries)	Food for Home		Meals Away		Clothing		Other	
		Mean	Median	Mean	Median	Mean	Median	Mean	Median
Baseline	2,402	19.6	16	4.1	3	0.9	0	8.5	7
Period 1	2,037	17.1	11	1.7	1	0.4	0	6.9	5
Period 2	600	12.7	5.5	1.9	1	0.5	0	6.0	4
Telephone	401	12.1	6	1.6	1	0.4	0	5.4	4
Online	199	13.9	4	2.3	1	0.9	0	7.2	6

These counts show a decline in the median entries over time, for all sections except clothing. In addition to the section counts presented above, we can focus on the change in counts for specific section combinations for which the pandemic was expected to have differential impacts (Table 9).

Table 9. Percent of decline in average diary entries by section reported in subsequent time periods relative to baseline time period

	N (diaries)	Food for Home	Meals Away and	
			Clothing	Other
Baseline	2,402	19.6 entries	5 entries	8.5 entries
Period 1	2,037	-12.8%	-58.0%	-18.8%
Period 2	600	-35.2%	-52.0%	-29.4%

From this table, we can see the declines in entries were likely due, at least in part, to changes in spending behavior prompted by COVID-19. Among the Diary sections most affected by COVID-19-prompted closures (e.g., fewer clothing purchases due to the increase in shuttered stores, fewer restaurant meal purchases), CUs reported 58.0 percent fewer entries in period 1 than in the baseline period. This change in pattern from the baseline period for Meals Away, Clothing, and Other expenditures provides evidence that some of the diary reporting declines reflected an actual curtailing of purchases. At the same time, there was a decline of 12.8 percent in total entries in the Food for Home section in which grocery purchases are collected. This is counter to what would be expected during a period when at-home food expenditures were noted to have increased and the decline in this section’s entries offers evidence for a

²⁶ Wilcoxon Rank-Sum two sample test; $Z = 2.98$, two-sided ($p = 0.03$).

²⁷ The Consumer Expenditure Diary consists of four sections: “Food and Drinks for Home Consumption” (or “Food for Home”), “Meals, Snacks, and Drinks Away from Home” (or “Meals Away”) which includes delivered or take-out food, “Clothing, Shoes, Jewelry, and Accessories” (or “Clothing”), and “All Other Products, Services, and Expenses” (or “Other”).

change in diary reporting behavior.²⁸ Further examination of this, presented below, reveals that the incidence of unitemized expenditures (e.g., “groceries” instead of the preferred itemized descriptions “milk,” “bread,” etc.) in the Food for Home section increased across time periods (Table 10).

Table 10. Share of diary weeks with ‘groceries’ as unitemized Food for Home section entry and associated costs

	N (diaries)	% 'groceries'	Median ('groceries')	Median (all other)
Baseline	2,402	6.0%	\$53.51	\$3.15
Period 1	2,037	11.2%	\$71.49	\$3.56
Period 2	600	12.5%	\$79.06	\$3.85
Telephone	401	10.7%	\$76.00	\$3.68
Online	199	16.1%	\$88.78	\$3.99

From the baseline period through period 2 the percentage of diaries containing “groceries” as a descriptor more than doubled in that section – from 6.0 percent to 12.5 percent of diaries. Furthermore, the median for respondent-reported unitemized expenses for “groceries” increased over the time periods while the medians for itemized expenses remained in a range of \$3 to \$4.²⁹ This pattern suggests that some or all of the reduction in entries in Table 9 was associated with poorer reporting behavior among respondents who may have incurred unusually large grocery expenses in these later time periods. Additionally, this general “groceries” entry occurred more in online diaries than in phone transcription. In summary, then, changes in both spending and reporting behavior were likely contributors to the decline in expenditure counts.

Recalled Expenses

The number of diary entries is composed of those expenditures respondents reported throughout the week as well as forgotten expenditures that interviewers are instructed to collect (or prompt respondents to enter) through ‘recall’ at the end of the reporting period. Recalled expenditures are normally entered into the respondent’s diary at pickup, but after the suspension of personal visits, they were collected as part of the phone transcription process by the interviewer. In the case of CUs placed with online diaries, interviewers were instructed to enter recalled expenses in a separate paper diary, especially if the CU decided to discontinue online diary entry. Interviewers were asked to record among their notes about the interview whether CUs: 1) provided any entries in the diary, and 2) had any entries to be collected by recall. In this analysis, a diary with no entries except for those collected by recall was defined as a “total recall” diary, and one that had initial and recalled-entries was a “partial recall” diary. Our analysis also verified the interviewer reports of recall against the actual number of entries contained in diaries and found only minor inconsistencies. There was a range of 0.5 percent to 1.5 percent of diaries that involved interviewer reports of entries (initially or by recall) without any entries actually captured. This error increased over time periods, but occurred with similar frequency by mode. An examination of the rate of interviewer-reported recall is shown below (Table 11).

²⁸ See USDA Economic Research Service’s [“COVID-19 Economic Implications for Agriculture, Food, and Rural America.”](#)

²⁹ Medians are reported due to the potential for outliers at this stage of CE processing; some additional unitemized entries (e.g., ‘grocery receipt,’ ‘food,’ ‘grocerys’ [sic]) are also among those categorized as ‘all other.’

Table 11. Percent of diaries with total and partial recall³⁰ by time period and mode

	Any Recall	Total Recall	Partial Recall
Baseline (n=2,446)	17.2%	7.1%	10.1%
Period 1 (n=2,072)	30.6%	11.7%	19.0%
Period 2 (n=622)	20.7%	9.2%	11.6%
Telephone (n=418)	25.8%	11.7%	14.1%
Online (n=204)	10.3%	3.9%	6.4%

The table indicates that recall initially became more prevalent after the suspension of personal visits. For context, we examined the share of diaries involving data collected mainly from telephone (Table 12).

Table 12. Percent of diaries by mode in which majority of data collected (excluding missing)

	Mainly telephone	Mainly in-person
Baseline (n=2,426)	5.2%	94.8%
Period 1 (n=2,050)	89.8%	10.2%
Period 2 (n=610)	98.2%	1.8%
Telephone (n=412)	97.8%	2.2%
Online (n=198)	99.0%	1.0%

In the baseline period, 17.2 percent of diaries for which information was provided contained at least some recalled entries. In this time period, only 5.2 percent of diaries involved a majority of data collection by phone (when excluding missing values for collection mode). The rate of any recall increased to 30.6 percent of diaries in the period immediately following the suspension of personal visits, a time period when 89.8 percent of diaries were mainly collected by phone. Some interviewers may have classified the standard expenses recorded via phone transcription as recalled expenses. The lower rate of recall in period 2, in which 98.2 percent of diaries involved mainly phone collection,³¹ was partially a result of less recall among CUs placed with online diaries (with a rate of recall of only 10.3 percent and only 3.9 percent of diaries involving total recall).³² The prevalence of recall among online diaries was roughly half of that for a prior test of this mode (although that test differed procedurally in involving a dedicated location for interviewers to supply recalled expenses). We further examined the specific expenditures recorded for the 21 online diaries containing at least some recalled entries (the 10.3 percent of all diaries that were placed online). For online diaries, interviewers were instructed to enter the expenses respondents recalled in a paper diary. Examining the mode associated with entries, it was observed that 12 of the diaries had some or all of their entries entered by paper form,³³ with the remaining 9 having only online entries.

³⁰ This includes blank diaries but excludes diaries where response to the note about recall was blank or missing.

³¹ Online collection was not tracked as part of the variable (TELPV) used for Table 12 data. There were only 6 CUs (11 diaries) in this period in which personal visits occurred; for the others the mode of data collection was undetermined.

³² Online CUs with recall had a near-average number of entries for CUs in this time period, after including recalled entries, averaging 27 and 14 for the diaries with partial and total recall, respectively.

³³ Some of these diaries may have been filled out by respondents who were placed with an online diary, but changed their mind and proceeded to report their diary expenses by ‘paper form,’ as collected through phone transcription.

4. Drop-Off in Weekly Reporting Behavior

As an indicator of data quality, drop-off in reporting of expenditures suggests respondent fatigue with the diary reporting process over a two-week period, as there is no reason to expect one week would have more expenditures than another on average. As measured – the proportion of two-week diary expenditure entries reported in week 1 – any percentage higher than 50 percent represents drop-off. To adjust for CUs that may have had expenditures in one week but not the other (“blank diaries”) we limited analysis to CUs having at least one expenditure in both weeks. Table 13 shows both the average entries per week, and the computed average within-CU drop-off.

Table 13. Weekly entries and within-CU drop-off in weekly reporting (excluding blank diaries)

	Baseline (n=1,165)	Period 1 (n=972)	Period 2 (n=285)
Week 1 mean entries	35.3	28.0	22.7
Week 2 mean entries	31.8	25.5	20.4
Mean entries (total)	67.1	53.6	43.1
Mean (CU-level) drop-off	53.0%	52.4%	53.6%

This table reveals a very stable rate of drop-off across time periods, suggesting reporting behavior and fatigue was not affected by the changes in data collection protocol or the pandemic. The differences from one time period to the next were not statistically significant.³⁴ An examination of respondent drop-off (and increase) in supplying *any* entries in a diary from one week to the next is shown in Table 14 below.

Table 14. Extent of CU weekly participation by time period

	Both non-blank weeks		1st week only		2nd week only		Both weeks blank	
	CUs	%	CUs	%	CUs	%	CUs	%
Baseline	1,268	91.9%	40	3.2%	32	2.5%	31	2.4%
Period 1	1,074	90.5%	60	5.6%	33	3.1%	9	<0.1%
Period 2	318	89.6%	15	4.7%	15	4.7%	3	0.1%

There was a slight increase in CUs that only had entries in the first of the two reporting weeks from the baseline to period 1 (3.2 percent to 5.6 percent respectively), but this decreased slightly by period 2 (4.7 percent). Fewer CUs started recording entries after not doing so the first week, and a small number of the CUs being analyzed also had no entries.

5. Item Non-Response

Item non-response comes from inability or refusal to provide a valid value for one of the collected fields of the diary. The rate of item non-response for providing a cost amount is examined in Table 15.

³⁴ Wilcoxon Rank-Sum two sample test; Baseline to Period 1, $Z = -0.37$, two-sided ($p = 0.71$). Period 1 to Period 2, $Z = 1.02$, two-sided ($p = 0.31$).

Table 15. Item non-response in the cost field by time period and mode (excluding blank diary weeks)

	N (entries)	Item non-response
Baseline	79,541	0.52%
Period 1	53,322	2.07%
Period 2	12,505	1.10%
Telephone	7,816	1.61%
Online	4,828	0.23%

Out of all entries, only a half of a percent involved missing cost information during the initial baseline period we examined. This low rate increased slightly in period 1, but declined thereafter.³⁵ The decline in missing cost data was in part attributable to the improved collection of this among online diary CUs, in which only 0.23 percent of entries were blank.³⁶ There was a higher prevalence of missing cost information in the Food for Home section among CUs involving telephone transcription than among CUs with online diaries.

For item description, the Meals Away section does not involve respondents entering a description beyond providing the name of the vendor, and so this section was excluded from calculations (along with blank diary weeks). In Table 16, missing data were designated by the Census-supplied entry ‘BLANK’ (but they do not include the few entries that were illegible).

Table 16. Item non-response in the item description field by time period and mode (excluding blank diary weeks)

	N (entries)	Item non-response
Baseline	69,809	0.06%
Period 1	49,815	0.36%
Period 2	11,530	0.41%
Telephone	7,164	0.27%
Online	4,366	0.64%

As with cost, there tended to be very low rates of item non-response for the item description field. This did significantly increase from 0.06 percent to 0.36 percent across the initial time periods examined before leveling off.³⁷ Unlike with cost, there was a larger share of online entries missing an item description in period 2.³⁸

³⁵ Wilcoxon Rank-Sum two sample test; Baseline to Period 1, $Z = 27.04$, two-sided ($p < 0.0001$). Period 1 to Period 2, $Z = -7.04$, two-sided ($p < 0.0001$).

³⁶ Wilcoxon Rank-Sum two sample test; Telephone and Online, $Z = -7.30$, two-sided ($p < 0.0001$). In online diaries, respondents must fill out the business name, expenditure category and date fields, as well as one of the other fields. As a result, cost and description do not both need to be provided in order for the entry to be saved.

³⁷ Wilcoxon Rank-Sum two sample test; Baseline to Period 1, $Z = 11.63$, two-sided ($p < 0.0001$). Period 1 to Period 2, $Z = 0.77$, two-sided ($p = 0.44$).

³⁸ Wilcoxon Rank-Sum two sample test; Telephone and Online, $Z = 3.07$, two-sided ($p = 0.002$).

IV. Conclusion

This report found a significant decline in weekly expenditure entries in diaries across the time periods examined. Deeper analysis suggests some of this was attributable to actual changes in spending patterns, as there was a 58.0 percent reduction from the baseline period to period 1 in diary entries for Clothing and Meals Away sections, compared to only a 12.8 percent decline for Food for Home purchases. The decline in the reported number of Food for Home entries was unlikely to be reflecting an actual decline in entries for these CUs, as examination of item descriptions indicated that respondents may not have itemized grocery expenditures during a period in which they may have had unusually large grocery purchases; unitemized reports of “groceries” increased from 6.0 percent to 12.5 percent during these time periods. Time periods 1 and 2 involved a shift in the sample composition to include a larger proportion of homeowners and college graduates than in the baseline period, which also may have affected the spending reported. The unplanned introduction of the online diary as part of routine CE diary data collection was necessitated by the COVID-19 pandemic, but did not appear to negatively affect the quality of data received. Encouragingly, the extent of entries among CUs placed with online diaries was no less than those collected via phone transcription. We must concede that it is hard to disentangle the effects of protocol from the actual changes in spending patterns during this time period using these data. We explored the prevalence of respondent diaries indicating an absence of spending and found minimal differences over time. We plan to produce a subsequent report that will further analyze many of the changes examined in this report and include findings related to changes in the expenditure amounts reported.

Our examination of recalled expenditures suggested that changes in protocol or reporting behavior associated with COVID-19 were associated with an increasing number of CUs involving recall. This was attenuated with the introduction of online diaries, in which recall was carried out less frequently than in diaries involving phone transcription. Across these time periods, we found interviewer indication of entries/recall to be relatively accurate in association with the presence of entries. That said, the lower extent of recall occurring among online diarists may have caused some underreporting.

Respondent willingness to participate in diary keeping across weeks remained steady despite the changes in protocols and the impact of COVID-19. The share of entries provided in the two-week period was only slightly higher in the first week, although there was a slight increase in the percent of CUs that only completed one of the two diary weeks. Generally, data quality as measured by missing data was good throughout the time periods examined, with cost and item description fields missing data for approximately 2 percent of the entries, at most. Rates of missing data suggested this was most an issue immediately after the onset of the COVID-19 pandemic as it leveled off or declined in period 2.

Overall, the COVID-19 pandemic and associated procedural and spending changes had a negative effect on participation in the CE survey, changing the demographic characteristics of those who participated, and reducing the number of entries entered by those participating CUs. Further work will be necessary to investigate the extent to which differences in the collected data resulted from procedural changes and how much these differences were due to changes in how well respondents reported the expenditures they had.

Methods Appendix

We used unweighted data for this report, as they intended to refer only to the universe of complete diaries presented in Table 2, and not to the larger population.

To carry out the analysis, we used SAS and Stata software in addition to common computer and calculator software. We employed a mix of tests for significance depending on the analysis. For demographic analysis, we used a test of the equality of proportions.

For analysis of data quality, we used the Wilcoxon Rank-Sum two sample test. The Wilcoxon test is a non-parametric alternative to the independent samples t-test that can be used for data that is not normally distributed.