Large Scale Feasibility Test Final Report

December 6, 2021

Graham Jones, Parvati Krishnamurty, Brett McBride

Consumer Expenditure Surveys Program Report Series

1 EXECUTIVE SUMMARY

The Large-Scale Feasibility Test of the Online Diary (LSF) was administered between October 2019 and March 2020 with a close out of March cases in April 2020. The LSF included both online and paper diaries. A preliminary report using unprocessed data from October 2019 through February 2020 was completed in April 2021, which covered sample performance, data collection issues, debriefing results, preliminary findings for expenditure estimates and demographics. The purpose of this final report is to provide final response rates, final demographics, and conduct a more detailed analysis of expenditure counts and amounts using data that have been post-processed at BLS (including editing, allocation, imputation, and creation of weights). This report will also provide a detailed recommendation for implementation into the Consumer Expenditure Surveys. The executive summary highlights the major findings based on the analysis of the processed LSF data. Further detail on the findings and recommendations can be found in the report.

Response rates

- Final response rates were lower in the LSF compared to Production. Using the American Association for Public Opinion Research (AAPOR) RR2 definition, the overall LSF final response rate at the diary level[1] for October 2019 through March 2020 was 41.75 percent which was lower than Production (46 percent) over the same six months. Historically, response rates for field tests at Census have been lower than for the Production survey.
- The proportion of refusals were comparable between LSF and Production, and the rate of noncontacts was slightly lower in the LSF compared to Production.
- The difference in response rates was due to "Other" type A non-response outcomes, which were higher in the LSF. Part of the difference in the other non-response category may be attributable to a test effect in which diaries were not placed on time and the disposition was automatically coded as a non-response.

[1] Final CE response rates reflect estimation response rates and are calculated considering each week of the diary as a separate, independent interview.

- The remaining part of the difference in response rates is driven by a difference in the number of diaries removed due to suspected low quality during processing. Minimum expenditure edit rates were higher in the LSF compared to Diary Survey Production.

Sample characteristics

- LSF completed cases had a significantly higher proportion of non-urban consumer units (CUs) and single person CUs compared to Diary Survey Production. There were also small but non-significant differences by race, education level, age, income and homeownership. These higher rates could be due to the underlying sample draw being different for the LSF compared to Production. Another possibility is that the introduction of the new online mode type may have bolstered participation for these demographic sub-groups in the LSF.
- Comparing the demographics of online and paper diaries within the LSF, reference persons and CUs that used online diaries trended younger, were more likely to have attended at least "Some College", had higher average and median incomes, and were more likely to be non-urban CUs than those with paper diaries. Online CUs also were also less likely to have a reference person of Hispanic origin or be a single member CU.

Expenditure counts

- After processing, the median count of entries for Production CUs (31.0) was higher than that for LSF CUs (28.0), with a difference in medians of 3 entries. The mean number of entries was 36.5 entries for Production and 35.5 entries for LSF CUs.

Expenditure amounts

- After processing was complete, weighted total expenditures for Production were higher than the LSF test. The $379.50 median amount for Production was $37.37 larger than that for LSF ($342.13), a 10.4 percent difference. The average total for Production was $726.01 compared to $629.30 for the LSF, a significant difference of $96.71 (a 14.3 percent difference).
- The difference was driven almost entirely by lower reporting in the "Other" expenditure category. There was a significant mean difference of $91.80 between the amounts spent for the Production group ($531.97) and the LSF group ($440.17) for the "Other" section

with a difference in medians of $38.88. Isolating items that are sourced from the Diary for publications, however, the differences in the "Other" category were no longer significant. There was no significant difference in expenditures for the Food at Home or Clothing categories and only a minor, but significant difference in Food Away From Home.

- Within the LSF group, after Edit and Estimation System processing, online diarists recorded a higher median weekly expenditure of $389.55 compared to paper diarists with a median total of $333.58.
- Regression analysis found that online diary respondents reported significantly lower diary expenditure totals than production paper diarists, even after controlling for demographic and geographic correlates.

Minimum Expenditure Edit and Allocation rates

- Minimum expenditure edit rates were higher in the LSF compared to Production (5.9 percent compared to 3.8 percent).
- Allocation rates were higher in the LSF compared to Production (12.1 percent vs. 9.3 percent). In the LSF, allocation rates were also higher for the online compared to the paper diary mode. This indicates higher rates of bundling (e.g., writing "groceries" rather than itemizing expenses at a grocery store) in the online diary.
- Online diaries had higher rates of allocation across all categories in the LSF. The largest difference between online and paper allocation rates was in the "Other" expenditure category (17.2 percent versus 8.8 percent respectively), followed by "Clothing" (19.1 percent versus 12.3 percent), then "Food for Home" (15.9 percent versus 9.9 percent) and finally "Meals Away from Home" (13.5 percent versus 12.5 percent). Meals Away from Home had the smallest difference, likely due to relatively less need for itemizing Meals Away from Home expenditures, which are less likely to be part of larger shopping trips.

Using Data Quality Assessment to Target Online Mode

- Field Representatives (FRs) were asked to assess the data quality of the diaries in the LSF. They reported that 25 percent of diaries were of low data quality. However, it is not known how FRs assessed data quality. Our results using FR data quality assessments did not agree with quality assessments using data methods such as allocation rates.

- A comparison of the FRs quality assessment for online and paper diaries in the LSF indicates that online diaries had higher data quality than paper. Diaries were of very high or high quality in 91.1 percent of online diary cases, compared to 72.7 percent of paper diary cases.
- Education level, race, and income were significant predictors of FR reported online diary quality. Diaries with an Asian respondent were rated to be of lower quality compared to those of other races, however the Asian sample was very small. Respondents with a high school education or less had significantly lower quality diaries. FR reported diary quality was also found to be lower for CUs with lower imputed income levels, however we do not have this information prior to diary placement to help with assigning optimal mode. We recommend further investigation into the possibility of using education level as information of targeting households for online placement.

Recommendations for implementation

Despite findings of lower expenditure reporting in the LSF, we recommend proceeding with an official production implementation in FY22 due to the ongoing need for an online data collection instrument which addresses respondent and field staff concerns about social distancing, positive feedback from data collection staff, requests from respondents, and specific steps to mitigate identified drivers of lower expenditure reporting. Figure 1.1 outlines our recommendations for further research and improvements to the instrument and protocols:

Figure 1.1. Recommendations

1. Explore performance-based incentives and other ways to improve data quality.
2. Provide better help desk access and training for help desk staff to help with login issues.
3. Make sure as many FRs as possible login with the respondent during the pickup visit.
4. Provide a way for FRs to edit the online diary data in the instrument similar to how they would edit a paper diary.
5. Allow FRs to disclose that they are able to view the respondent's entries to facilitate a conversation of encouraging higher data quality.

6. Explore adding features to the online diary instrument such as instructions and examples for "Other" category in the online diary. Consider adding categories to split the "Other" into more easily understood categories.
7. Do not close out the instrument at pickup to allow FRs more time to enter expenses from receipts.
8. Investigate online placement procedures in more detail to see whether to exclude online placement for certain categories such as those with an education level of high school or less as this group was identified as having lower data quality in the LSF.

2 OVERVIEW

2.1 BACKGROUND

In 2013, the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CE) approved a survey redesign plan with the objective to improve the quality of the survey estimates through a verifiable reduction in measurement error. BLS will be realizing the CE survey redesign through a phased implementation of key design elements. This involves implementing an online, CU-level diary into CE Diary Survey Production in July 2022 (pending results and recommendations from the Large Scale Feasibility test described here) and implementing a streamlined questionnaire into CE Interview Survey production April 2023.

In order to assess potential fielding issues and evaluate data quality of online diaries, the CE program completed the LSF. The LSF was built on information learned from prior online diary tests and was planned to have sufficient sample to make statistical inferences. An online diary was developed at Census based on requirements from an online diary designed by Westat[2]. The LSF was fielded between October 2019 and April 2020. A preliminary report using unprocessed data from October 2019 through February 2020 was completed in April 2021, which covered sample performance, data collection issues, debriefing results, preliminary findings for expenditure estimates and demographics. The purpose of this final report is to provide final

[2] The Westat report is available at https://www.bls.gov/cex/research_papers/pdf/ce-online-diary-usability-testing.pdf

response rates, final demographics, and conduct a more detailed analysis of expenditure counts and amounts using data that has gone through post-processing at BLS (including editing, allocation, imputation, and creation of weights). This report will also provide a detailed recommendation for implementation into the Consumer Expenditure Surveys.

3 PARTICIPATION

3.1 SAMPLE DESCRIPTION

The LSF starting sample frame consisted of 2,502 housing units selected from the CE Diary reserve sample. The sample included 2,478 unit frame sample units and 24 group quarters frame units. The sample covered the continental United States, excluding Hawaii and Alaska. The sample was fielded with placements occurring over 6 months (October 2019-March 2020). The final close out of the test was at the end of April.

The CE Diary Survey Production sample, which makes up the control group for this project, included a total of 2,962 CUs that provided complete diaries between October 2019 and March 2020 (Figure 3.1.1) [3]. In the same time frame, 901 CUs in the treatment group (LSF) had complete outcome codes (Figure 3.1.2).

Figure 3.1.1. Production diary-level sample totals[4]

Cases (CUs) (complete)	*Both-week diary CUs*	*One-week diary CUs*	*No-entry CUs*
2,962	2,767 (93.4%)	76 (2.6%)	119 (4%)

Diary totals	*Non-blank diaries*
5,838	5,618 (96.2%)

Figure 3.1.2. LSF diary-level sample totals

[3] Note that sample sizes increased in January 2020 (12,000 annual production addresses to 17,800 addresses).
[4] To derive diary counts for "no-entry" and "blank" diaries, we examined the expenditure total variable (ZTOTAL), at the case and diary level. Those with a ZTOTAL of zero were deemed "no entry" or "blank" diaries.

Cases (CUs) (complete)	*Both-week CUs*	*One-week diary CUs*	*No-entry CUs*
901	842 (93.5%)	19 (2.1%)	40 (4.4%)

Diary totals	*Non-blank diaries*
1,782	1,701 (95.5%)

In Production, 2,767 of the 2,962 CUs completing diaries completed them in both weeks, a rate roughly equivalent to that of the LSF sample – 93.4 percent and 93.5 percent, respectively. When examining the number of diaries among complete CUs in the samples, there were 5,838 diaries (96.2 percent of which had an expenditure total greater than 0) from Production, and 1,782 diaries (95.5 percent with a non-zero expenditure total) from LSF.

3.2 SAMPLE CHARACTERISTICS

The demographic composition of the LSF consumer units (CU's) was similar to the make-up of their Diary Production sample counterparts and no statistical differences were found among the categories. A complete breakdown of the demographic composition of the LSF sample and the Diary Production sample are illustrated in Figure 3.2.1 below.

Figure 3.2.1 Sample characteristics of LSF Completed Cases

Demographic category	*LSF Complete cases*	*Production Diary Complete cases*	*Percent Difference (LSF minus Production)*
No. of Completed Cases	1,782	5,838	
Race of Reference Person			
White	78.0	79.7	-1.7
Black	13.2	12.8	0.4
Other (incl. Asian, multi, other)	8.8	7.6	1.2
Hispanic Origin of Reference Person			
Hispanic	15.2	14.0	1.2
Non-Hispanic	84.8	86.0	-1.2
Area Type			

Demographic category	*LSF Complete cases*	*Production Diary Complete cases*	*Percent Difference (LSF minus Production)*
Urban	90.4	92.4	2.0
Non-Urban	9.6	7.6	-2.0
Age of Reference Person			
Under 25 years	4.6	5.4	-0.7
25-34 years	17.7	16.6	1.1
35-49 years	24.4	24.7	-0.3
50-64 years	27.5	27.6	-0.1
65 years and older	25.8	25.7	0.1
Education of Reference Person			
Less than high school	10.2	8.2	2.0
High school graduate	22.8	21.1	1.7
Some college	30.4	32.6	-2.2
College graduate	36.5	38.1	-1.6
CU Size			
Single person	31.6	29.7	1.8
2-3 persons	46.2	47.6	-1.4
4+ persons	22.3	22.7	-0.4
Mean Income Before Tax	$80,520	$83,412	-$2,892
Median Income Before Tax	$57,261	$59,998	-$2,737
Housing Tenure			
Renter	35.7	35.7	0.0
Owner	64.3	64.3	0.0
A weighted comparison of means found no statistical differences between samples, using $p < 0.05$			

3.3 DEMOGRAPHICS OF MODE CHOICE

In this section, we compare the respondents who chose the online diary mode for the LSF to those who were ineligible for the online diary or otherwise chose the paper diary mode.

Figure 3.3.1 presents the demographic composition of LSF CUs and reference persons by the mode type associated with the given complete diary. A reference person is defined as 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." Reference persons and CUs that used online diaries trended younger, were more likely to have attended at least “Some College”, had higher average and median incomes, and were more likely to be non-urban CUs than those with paper diaries.

Online CUs also were also less likely to have a reference person of Hispanic origin or be a single member CU.

The analysis shows that 57.5 percent of online diary cases had a reference person who was under the age of 50, compared to 39.4 percent of the paper diary cases. Additionally, 83.4 percent of the online group had at least "Some College" as their highest level of education within their CU, compared to only 57.4 percent of the paper group. As would be expected, average income differences followed the trend of education. Those CUs with online diaries earned a mean average before tax income of $102,243 versus $69,418 for the paper group. This trend persisted for the median before tax income as those CU's placed with an online diary had a median of $83,143 versus $48,782 for the paper group[5]. The analysis also showed that 70 percent of those using the online diary were homeowners compared to 61.3 percent of those using the paper diary. Lastly, the percentage of online diary cases whose reference person reported being of Hispanic origin was much lower than the paper diary (9.2 percent vs 18 percent).

[5] Median income by diary mode was tested using a weighted bivariate quantile regression.

Figure 3.3.1 Demographics by Mode of Completion (N=1,737)[6]

Demographic category	*LSF Online diary*	*LSF Paper diary*	*Percent Difference (Online minus Paper)*
No. of Consumer Units	663	1,074	
Race of Reference Person			
White	79.6	77.0	2.6
Black	11.4	14.5	-3.1
Other (incl. Asian, multi, other)	9.0	8.5	0.5
Hispanic Origin of Reference Person			
Hispanic*	9.2	18.0	-8.9
Non-Hispanic*	90.9	82.0	8.9
Area Type			
Urban*	88.4	93.9	5.4
Non-Urban*	11.5	6.2	-5.4
Age of Reference Person			
Under 25 years*	6.6	3.4	3.2
25-34 years	18.4	16.4	2.0
35-49 years*	32.5	19.6	12.9
50-64 years	26.5	29.1	-2.6
65 years and older*	16.0	31.5	-15.5
Education of Reference Person			
Less than high school*	2.7	14.6	-11.8
High school graduate*	13.9	28.1	-14.2
Some college*	36.2	27.1	9.1
College graduate*	47.2	30.3	16.9
CU Size			
Single person*	28.4	33.1	-4.7
2-3 persons	48.9	45.0	3.9
4+ persons	22.7	21.9	0.8
Mean Income Before Tax*	$102,243	$69,418	$ 32,825
Median Income Before Tax**	$81,659	$46,245	$ 35,414
Demographic category	*LSF Online diary*	*LSF Paper diary*	*Percent Difference (Online minus Paper)*
Housing Tenure			
Renter	30.0	38.7	-8.70
Owner	70.0	61.3	8.70
*Indicator of statistical significance from a weighted comparison of means between mode types, using $p < 0.05$ **Indicator of statistical significance from a weighted bivariate quantile regression, using $p < 0.05$			

[6] The number of total diaries for the mode of completion analysis in Figure 3.3.1 is lower than in the sample characteristics in Figure 3.2.1 due to 45 diaries being removed because of their '217' outcome status.

3.4 OVERALL FINAL RESPONSE RATES

For the LSF test, we calculated the overall response rates similar to the way the final response rates are calculated in CE Diary Production. Using the AAPOR RR2 definition[7] the response rate is calculated as the total number of completed diary weeks divided by the total number of eligible interviews also in diary weeks (the sum of interviews plus Type A non-interviews). Type A non-interviews are interviews that were not completed by respondents due to either respondent refusal or the inability to reach respondents. Cases that were identified as temporarily absent by the FR are counted as completes in the calculation of final response rates at BLS. Type B/C refers to various types of ineligible cases including vacant or abandoned housing units and businesses.

Figure 3.4.1 presents the overall LSF response rates compared to similarly calculated CE Diary Survey Production response rates for the same months of October 2019 through March 2020. The overall LSF response rate was lower than Production (41.75 percent vs. 46 percent, respectively). The rate of refusals is comparable between the LSF and Diary Survey (25.96 percent vs. 25.53 percent), and the noncontact rate (which includes cases when no one was home) is slightly lower in the LSF compared to Production (6.65 percent vs. 7.4 percent). The observed differences in the overall response rates seem to be driven by the difference in other type A rates which are higher in the LSF compared to Production (25.63 percent vs. 21.22 percent). The other type A nonresponse category includes outcome codes for type A other, week 2 diary picked up too early, diary placed too late, or diaries with the majority of entries made by recall and picked up after 15 days. When diaries are not placed by the end of the month, the instrument automatically codes these as diary placed too late. The rate of diary placed too late is higher in the LSF compared to Production, and this could reflect a test effect i.e., that FRs prioritized placing the Production diaries over the test (LSF) diaries.

[7] AAPOR RR2 definition can be found at "https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf"

Figure 3.4.1 Overall Response Rates Compared to Diary Survey Production[8]

Category	*LSF count*	*LSF rate*	*Diary Survey Production count*	*Diary Survey Production rate*	*Percentage difference (LSF-Prod)*
Eligible Diaries (Type A+ Completes)	4,268	100	12,690	100	
Completes	1,782	41.75	5,838	46.00	-4.25
Overall nonresponse	2,486	58.25	6,852	54.00	4.25
-Noncontact	284	6.65	939	7.40	-0.75
-Other type A	1,094	25.63	2,673	21.22	4.41
-Refusals (type A)	1,108	25.96	3,240	25.53	0.43

4 DATA QUALITY

This section investigates the comparative quality of the online diaries in terms of the number of entries recorded, expenditure totals, allocation rates, legitimate blank diaries, and comparisons with demographic controls. Other factors such as the respondent technological problems, diary placement issues, and uncertainty about the diary task could account for poorer data quality among groups, these issues are discussed at length in the preliminary report and therefore not discussed here.

4.1 OVERALL DIARY EXPENDITURE TOTALS AND COUNTS BY GROUP

We examined the expenditure totals recorded in diaries, applying final weights reflecting population totals and the sample design (Figure 4.1.1).

[8] Diary level response rates as calculated for CE Diary Survey using the AAPOR RR2 definition.

Figure 4.1.1. Weighted diary expenditure totals, by group

Totals	*Production*				*LSF*				*Diff. in Means (LSF-Prod)*	*% Diff.*
	Diaries	*Mean*	*Mean (SE)*	*Median*	*Diaries*	*Mean*	*Mean (SE)*	*Median*		
Total	5,838	$726.01	$24.42	$379.50	1,782	$629.30	$39.03	$342.13	-$96.71*	14.3%
*T-Test; Z = -2.10, two-sided (p=0.041).										

Figure 4.1.1 shows that production CUs recorded a larger total amount of diary expenditures than did LSF CUs. The $379.50 median amount for Production was $37.37 larger than that for LSF, and the average totals were $96.71 higher, representing a percent difference of 14.3[9]. Although these differences do not control for differences in the sample demographics, they do reflect both imputation and other data processing edits applied to published CE Diary Survey estimates.

We further investigated these differences by comparing weighted expenditure amounts recorded by diary section.

Figure 4.1.2. Weighted diary expenditure totals, by section and group

Totals	*Production (n=5,838)*		*LSF (n=1,782)*		*Difference in Means (LSF-Prod)*
	Mean	*Median*	*Mean*	*Median*	
Food at Home	$102.64	$74.51	$104.94	$73.09	$2.30
Food Away	$58.18	$28.08	$50.86	$25.07	-$7.32
Clothing	$33.22	$0.00	$33.32	$0.00	$0.10
Other	$531.97	$188.75	$440.17	$149.87	-$91.80*
* T-Test; Z = -2.18, two-sided (p=0.038).					

[9] T-Test; Z = -2.10, two-sided (p=0.041).

Figure 4.1.2 shows that for most of the sections there were only minor differences in reported expenditures. The exception involved a significant mean difference of $91.80 between the amounts spent for the Production group ($531.97) and the LSF group ($440.17) for the "Other" section with a difference in medians of $38.88[10]. We examined the "Other" weighted section totals by whether UCCs are sourced from the diary or the interview for the published tables (Figure 4.1.3).

Figure 4.1.3. Weighted "Other" section expenditure counts, by whether sourced from diary and group

Totals	*Production (n=5,838)*		*LSF (n=1,782)*		*Difference in Means (LSF-Prod)*
	Mean	*Median*	*Mean*	*Median*	
Sourced	$69.27	$24.98	$63.50	$20.66	-$5.77
Not Sourced	$462.69	$126.00	$376.67	$100.68	-$86.02

The table above (Figure 4.1.3) shows that for sourced UCCs within the "Other" section there were only small differences between LSF and Production. These differences in mean and median amounts for sourced UCCs for production were not significant[11]. Most of the differences seen in Figure 4.1.2 appear to be attributable to items that are not sourced from the diary. A small portion of the "Other" section difference is attributable to a greater proportion of CUs with $0 in section expenditures among the LSF group (after allocation) – 11.5 percent of the total versus 9.2 percent for Production. A detailed examination of only the high-dollar sourced UCCs from the "Other" section, and how they differed by group can be found in Appendix A. In the "Food Away from Home" section, there was a smaller, but significant, difference among groups in mean expenditure amounts, with the Production group reporting $58.18 versus $50.86 for LSF[12].

[10] T-Test; Z = -2.18, two-sided (p=0.038).
[11] T-Test; Z = -1.35, p>|t|= 0.1765
[12] T-Test; Z = -2.43, two-sided (p=0.017).

The post-processed data does not capture the specific number of entries respondents provided in diaries as does the original pre-processed data due to allocation. Nevertheless, we examined the counts resulting from all CE data processing in Figure 4.1.4 below.

Figure 4.1.4. Weighted diary expenditure counts, by group

Counts	*Production*				*LSF*				*Difference in Means (LSF-Prod)*	*% Diff.*
	Diaries	*Mean*	*Mean (SE)*	*Median*	*Diaries*	*Mean*	*Mean (SE)*	*Median*		
Total	5,838	36.5	0.6	31.0	1,782	35.5	1.2	28.0	1.0	2.8%

Examination of diary counts (post-allocation) as shown in Figure 4.1.4 indicated the median count of entries for Production CUs was 3 higher than that for LSF CUs, with a difference in means of 1 entry. The mean for Production (36.5 entries) was higher than for the LSF (35.5 entries), a percent difference of 2.8, however the difference was not significant.

To control for demographic differences between CUs participating in the LSF and Production groups, we carried out regression analysis.

Figure 4.1.5. Weighted regression of mode and covariates on total expenditures (n=7,458)

	Model 1		*Model 2*	
	Coefficient	*SE*	*Coefficient*	*SE*
Intercept	-302.580	(220.67)	-55.323	(215.59)
LSF online	-113.591*	(46.27)	-115.003*	(46.39)
LSF paper	-73.055	(68.04)	-73.079	(66.96)
Log(pre-tax income)	112.939***	(17.98)	90.637***	(20.18)
Non-homeowner CU	-195.685***	(52.61)	-163.447***	(40.09)
Education of Reference Person: HS graduate or less	-269.340***	(72.74)	-252.074**	(75.16)
Education of Reference Person: Some college	-214.474**	(74.85)	-211.608**	(76.08)

	Model 1		*Model 2*	
	Coefficient	*SE*	*Coefficient*	*SE*
Education of Reference Person: Associate degree	-185.780*	(77.95)	-174.878*	(76.72)
Education of Reference Person: Professional degree	105.623	(97.72)	112.424	(98.57)
CU size: Single person			-54.337	(39.64)
CU size: 3 persons			-21.898	(54.64)
CU size: 4 persons			166.904	(95.92)
CU size: 5 persons			94.544	(67.20)
Age of reference person: Under 25 years			-79.992	(69.52)
Age of reference person: 25-34 years			-3.107	(69.13)
Age of reference person: 35-44 years			30.103	(56.61)
Age of reference person: 45-54 years			117.433	(68.20)
Age of reference person: 55-64 years			68.443	(48.15)
Reference person male			-43.606	(45.81)
Reference person Black			-111.750	(99.55)
Reference person Hispanic			-82.477	(58.48)
New York RO			-76.835	(62.18)
Philadelphia RO			-31.918	(65.91)
Chicago RO			47.708	(73.52)
Atlanta RO			-73.015	(58.99)
Denver RO			8.534	(53.11)
Model R^2	0.035		0.040	

*p<0.05, **p<0.01, ***p<0.0001

The regression results shown in Figure 4.1.5 indicate that online diary respondents reported significantly lower diary expenditure totals than production paper diarists, even after controlling for demographic and geographic correlates. To account for missing cases ('temporarily absent') that would have affected analysis, the model analyzed the subset of complete CUs with non-missing expenditure total (ZTOTAL) values. The final model found that being placed with an online diary (relative to a production paper diary) was associated with approximately $114 less in reported expenditure totals, controlling for other factors. The model did not find a LSF paper diary to be associated with a significant difference in expenditure totals. Three other variables were found to be highly significantly associated with expenditure amounts – income (log transformed), housing tenure, and the education level of the CU's most educated member. For income, a one percent increase in pre-tax income was associated with a $1.13 increase in total

reported expenditure amount. For tenure, a non-homeowner CU was associated with $196 less in total reported expenditures. Finally, compared to CUs with a member obtaining a college degree, those with members having lower education levels were all associated with lower reported expenditure amounts. The full model did not find the CU's size, respondent's age, sex, race or ethnicity, or the RO in which the interview occurred to be significant factors in reported expenditures. The regression models account for very little of the variation in expenditures as noted by the low R^2 values.

4.2 EXPENDITURE TOTALS BY MODE WITHIN THE LSF SAMPLE

Participation in the LSF occurred through online or paper diaries, depending on eligibility[13]. The placement mode, or the participants selected into the online mode, appeared to affect weighted expenditures (Figure 4.2.1).

Figure 4.2.1. LSF weighted expenditure totals by mode (N=1,738)

Mode	*Diaries*	*Average*	*Median*
Paper	982	$609.02	$333.58
Online	756	$696.21	$389.55

As seen in Figure 4.2.1, after processing, online diarists recorded a median of $389.55 in weekly expenditures; with paper diarists having a median total of $333.58. The amounts recorded by section are shown below (Figure 4.2.2). Examination of the weighted section entries reveals that LSF online diarists tended to provide mean and median amounts at or above those of LSF paper diarists. The median section totals supplied by LSF online diarists exceeded those of Production diarists, with the exception of amounts for the "Other" section, where Production diarists reported $188.75 in expenditures and LSF online diarists reported $175.95. However, the regression analysis in the prior section found the larger LSF online amounts were largely a function of the demographic characteristics of those CUs.

[13] Note: There were 44 diaries that were 'temporarily absent.' These were not included in the mode comparisons.

Figure 4.2.2. LSF weighted expenditure totals by section and mode (N=1,738)

Totals	*LSF paper (n=982)*		*LSF online (n=758)*		*Difference in Means (Online - Paper)*
	Mean	Median	Mean	Median	
Food at Home	$95.91	$69.47	$123.78	$86.89	$27.87
Food Away	$46.97	$19.98	$59.33	$39.00	$12.36
Clothing	$27.80	$0.00	$42.90	$0.00	$15.10
Other	$438.33	$149.87	$470.19	$175.95	$31.86

4.3 MINIMUM EXPENDITURE EDIT AND ALLOCATION RATES

During processing, some diaries have their status changed to a nonresponse because they are suspected to not accurately reflect the respondent's actual spending. This process is referred to as the minimum expenditure edit[14]. Identified cases are reclassified from complete to Type A nonresponse due to diaries having zero expenditures in one or both weeks of the diary keeping period or expenditures below certain thresholds based on CU size. The rate of minimum expenditure edits in the LSF compared to production is shown in Figure 4.3.1 below. Minimum expenditure edit rates are higher for the LSF compared to Production (5.93 percent vs. 3.79 percent). The online diary had a checkbox where respondents could indicate whether they had no expenditures in one or both weeks. Unfortunately, data from this checkbox was not captured in the online instrument and could not be used for analysis.

Figure 4.3.1 Minimum Expenditure Edit Rates

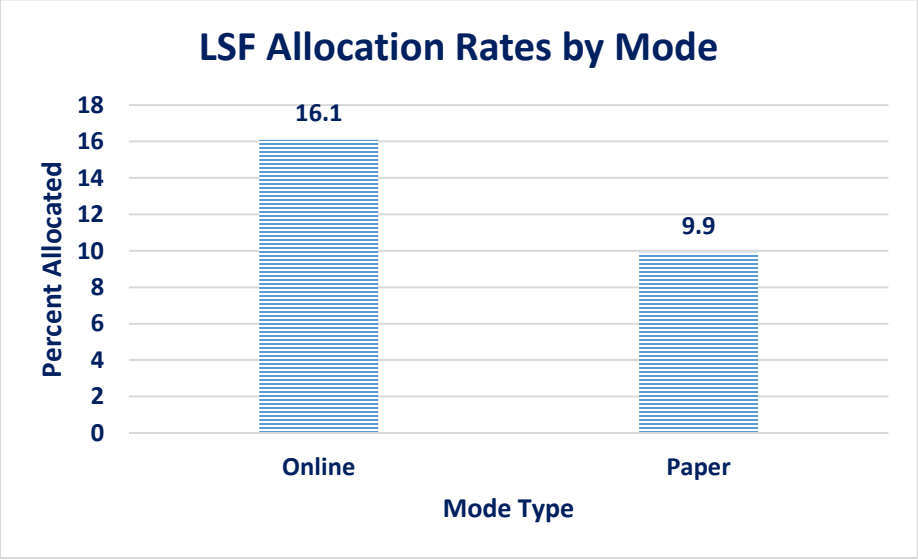
Type	*LSF count*	*LSF percent*	*Diary Survey Production count*	*Diary Survey Production percent*
Minimum expenditure edits	253	5.93	481	3.79

[14] More details on the minimal expenditure edit can be found in the Data Quality Profile Reference Guide under "Nonresponse reclassification"

Allocation edits are applied when respondents provide insufficient detail to meet tabulation requirements. For example, if a respondent provides a non-itemized expense amount with an item description of "food" or "groceries", that overall amount is allocated during processing to items in the Food at Home category. This is one of the key metrics used by CE to evaluate data quality. Generally, lower rates of allocation edits are preferable as this is associated with less chances of processing errors.

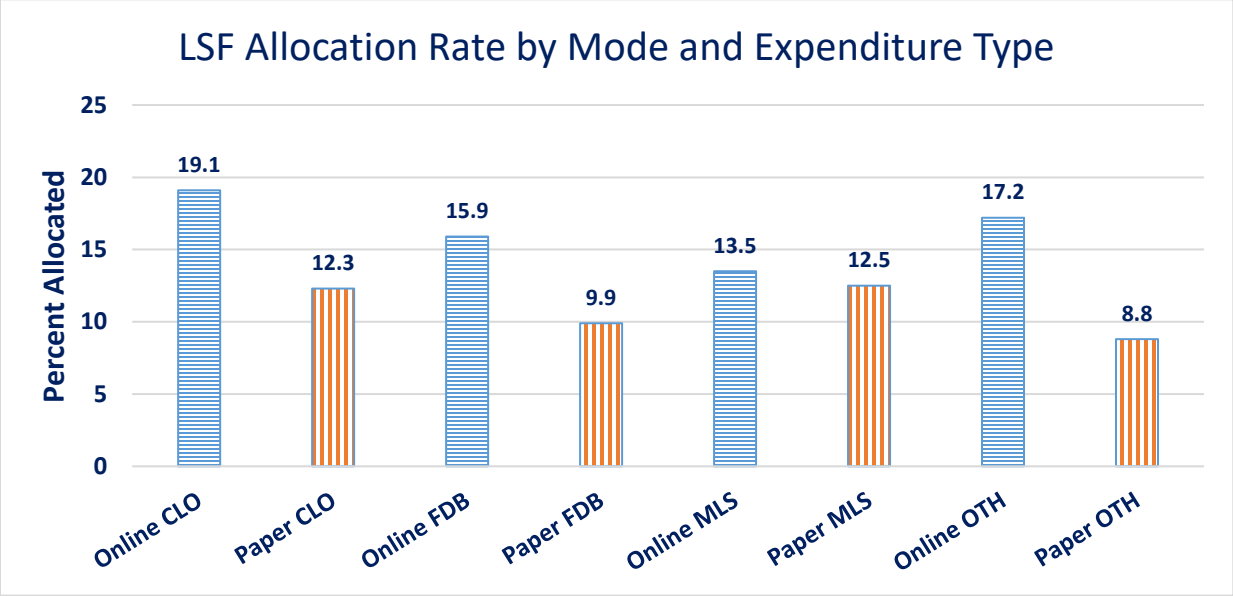
Overall allocation rates were higher in the LSF compared to Production (12.1 percent vs. 9.3 percent) indicating a greater degree of bundling of expenditures in the LSF. Looking at allocation rates by mode, allocation rates were higher in the LSF among online diaries (16.1 percent) compared to paper diaries (9.9 percent) as shown in Figure 4.3.2. The reason for this difference could be due to the different format of the online diary compared to the paper diary. The paper diary has many empty rows for respondents to enter expenses thus cueing them to itemize while the online diary just has one box at a time for each expenditure entry. Also, examples of the kind of items to be entered are more readily visible in the paper diary while respondents have to click on information buttons or help menus to see examples in the online diary. Additionally, in the production diary, an FR can correct expenditures in the paper diary that are not itemized during the pickup interview or afterwards, using receipts. It is unclear how this would be done in the online mode without the FR having access to the respondent's online diary.

Figure 4.3.2. LSF Allocation Rates by Mode of Completion



When looking at allocation rates by mode and expenditure type in Figure 4.3.3, we find that on average, online diaries had higher rates of allocation across all categories. The most prominent difference between online and paper allocation rates was in the "Other" expenditure category (17.2 percent versus 8.8 percent respectively), followed by "Clothing" (19.1 percent versus 12.3 percent), then "Food for Home" (15.9 percent versus 9.9 percent) and finally "Meals Away from Home" (13.5 percent versus 12.5 percent). In general, the difference in allocation rates were fairly steady across expenditure types for online and paper diaries, with the exception of Meals Away from Home, which only differed by 1 percentage point. This is likely due to relatively less need for itemizing Meals Away from Home expenditures compared to the other expenditure types, which are more likely to be part of a larger shopping trip.

Figure 4.3.3. LSF Allocation Rates by Mode of Completion and Expenditure Type



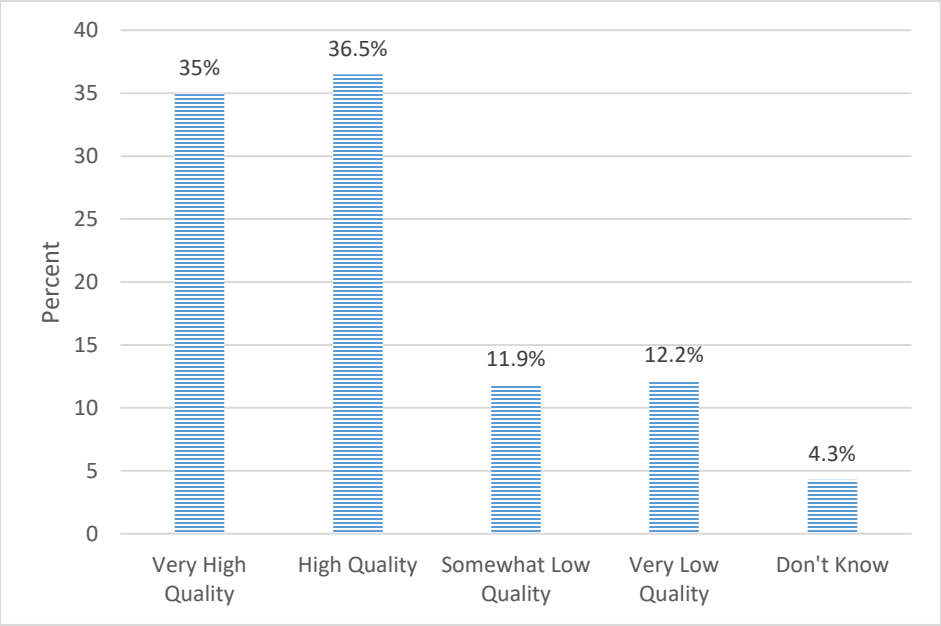
4.4 USING DATA QUALITY ASSESSMENT TO TARGET ONLINE MODE

4.4.1 Data Quality Assessment

FRs were asked to assess the data quality of the diaries, though we do not really know what heuristic FRs used to rate the diaries. The data quality assessment is shown in Figure 4.4.1. FRs

reported high or very high data quality in about 75 percent of cases where they rated the data quality. However, about 25 percent reported that the diaries were not high quality. The most frequent explanation provided was respondent disinclination to fill the diary including being busy, disinterested or concerns with privacy. The second most frequent explanation was respondent limitations such as mental, physical or language barriers. There were also some cases of the respondent not following instructions (such as itemizing groceries) and technical limitations (such as technical abilities, computer, or login issues).

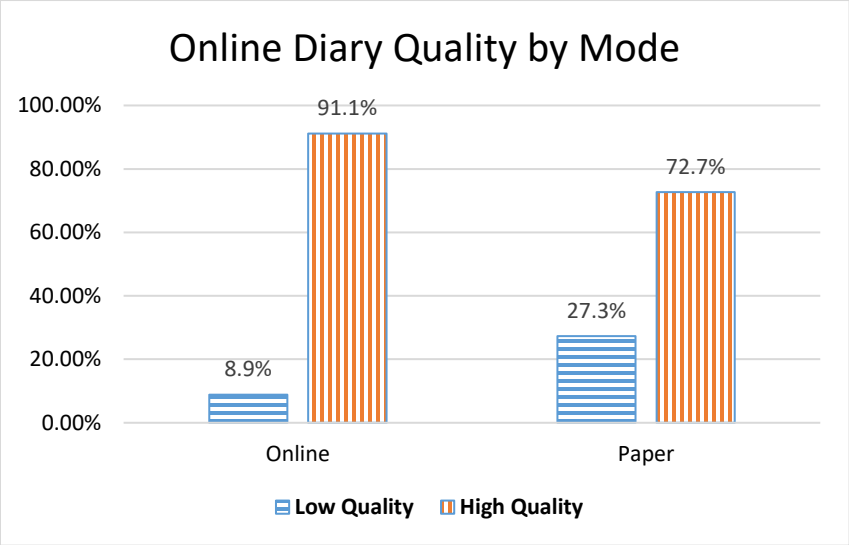
Figure 4.4.1. FR perception of data quality of LSF diaries (n=879)



4.4.2. Diary Quality by Mode of Completion

Figure 4.4.2 below shows the results of the FRs quality assessment for online compared to paper diaries in the LSF. This examination of the FR reported quality metric reveals that diaries were of very high or high quality in 91.1 percent of online diary cases, compared to 72.7 percent of paper diary cases.

Figure 4.4.2 Diary Quality by Mode of Completion (n=1,506)



For the 8.87 percent of online diaries that received a "Low" or "Very Low" quality rating, the most commonly cited reason was due to the CU being too busy during the diary keeping period, or not providing sufficient detail in their diary entries. The exact heuristics used by FRs to judge diary quality is uncertain, but on average FRs rated online diaries as high quality a higher percentage of the time than paper diaries.

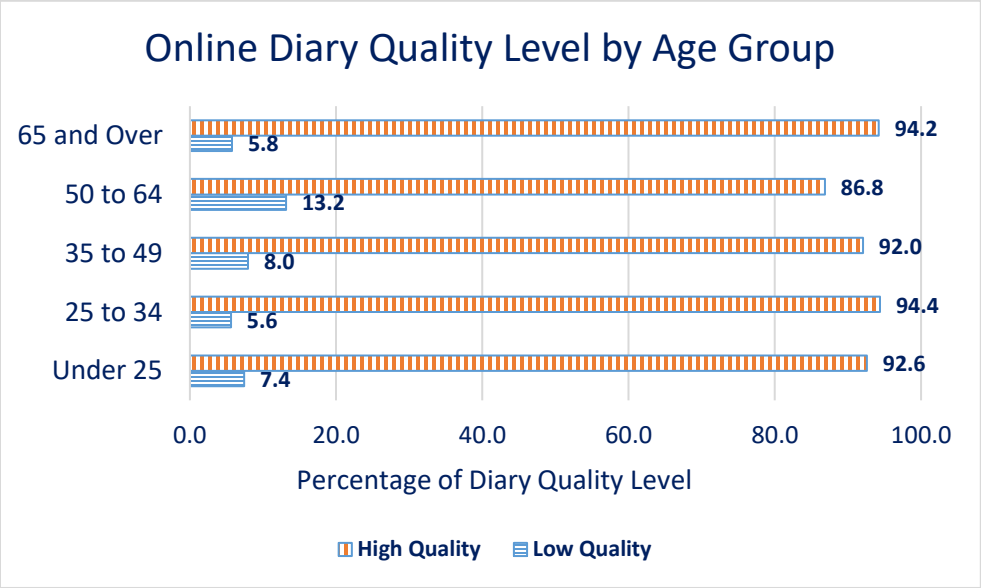
4.4.3 Online Diary Quality by Demographic Groups

Further analysis of the FR debriefing question on LSF diary quality, revealed that reported quality for online diary users varied significantly between certain demographic groups. Age, education, race and income, all initially appeared to have significant variation in reported quality within their respective groups.

When examining reported diary quality by age, we found that the percentage of online diaries reported to be of "Very High Quality" was the highest for reference persons under the age of 25, at 74 percent. All other age groups had lower rates of "Very High Quality" diaries, but higher rates of online diaries reported to be of "High Quality". This mixed result motivated the decision to collapse "Very High Quality" with "High Quality", and "Very Low Quality" with "Low Quality" to generate a binary indicator of FR reported data quality. Figure 4.4.3.1 shows that after this change the relationship between age and quality was no longer significant. Cases where

the reference person was in the "50 to 64" age group had a noticeably lower rate of high quality online diaries than the other groups, but the difference was not significant.[15]

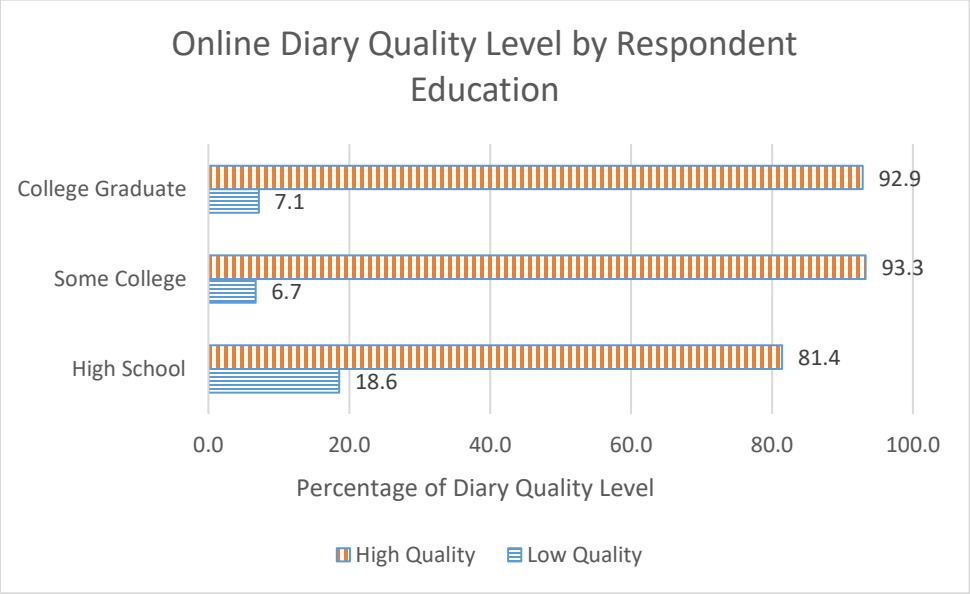
Figure 4.4.3.1 Online Diary Quality by Age Group (n=586)



An examination of reported online diary quality by education level of the respondent is shown in Figure 4.4.3.2 below. Online respondents who had some college experience or a college degree were associated with high quality diaries roughly 93 percent of the time for both groups. Meanwhile, respondents with a high school education or less had a significantly lower rates of high quality diaries at about 81 percent. We also looked at online diary quality by the education level of the reference person in the CU and the results were similar.

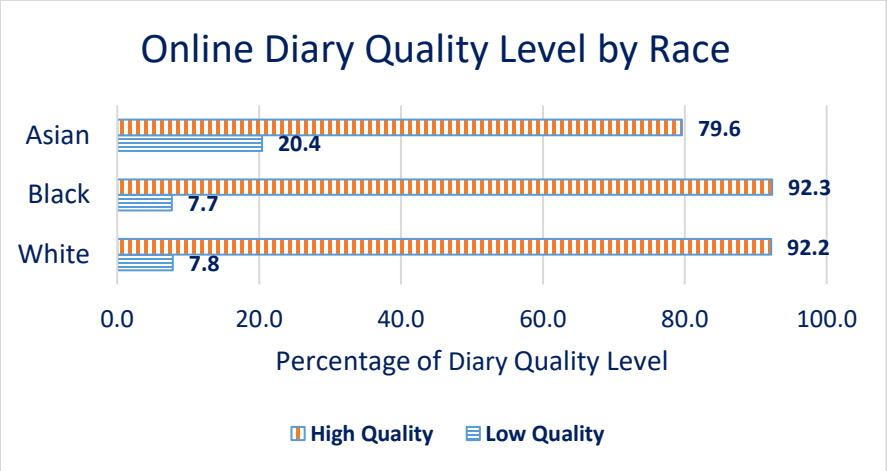
[15] It should be noted that of the 663 completed online diaries, only 586 were assigned a quality rating by the FR.

Figure 4.4.3.3. Online Diary Quality by Respondent Education Level (n=586)



Reported quality of online diaries by race of reference person, below, also showed statistically significant variation. Reference persons whose self-identified race was either "Black" or "White", showed nearly identical rates of high quality and online diaries, with both group percentages roughly equaling 92 percent. Reference persons whose self-identified race was "Asian" had a significantly lower rate of high quality online diaries at roughly 80 percent.[16]

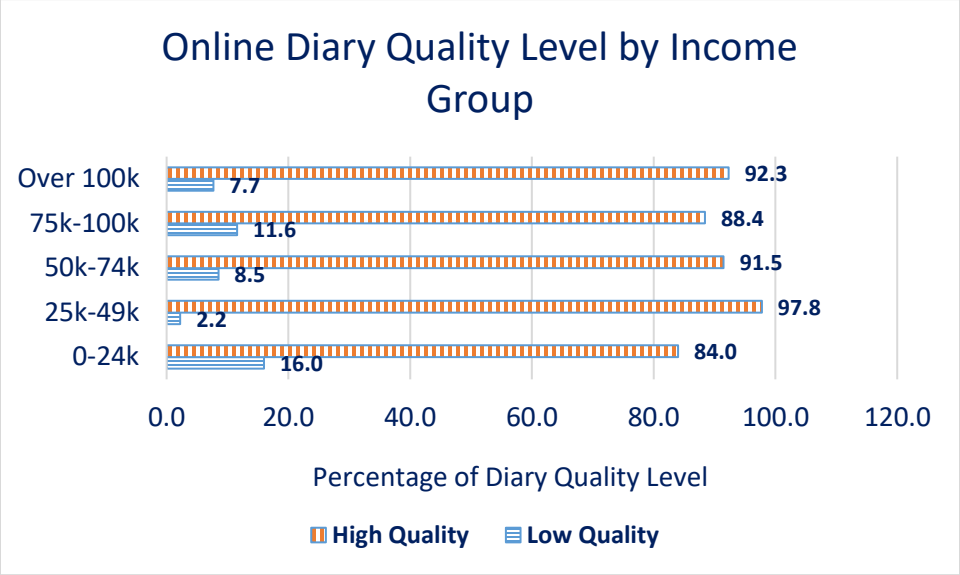
Figure 4.4.3.4. Online Diary Quality by Race of Reference Person (n=586)



[16] It should be noted that the number of reference persons in online diary cases self-identifying as Asian was very low, and this relationship was not significant for paper diary users.

The relationship between FR reported diary quality and CU income level also showed to be statistically significant (Figure 4.4.3.5), but since this comes from the imputed incomes, it is not information that we would have access to prior to diary placement, and therefore would not help with assigning optimal mode.

Figure 4.4.3.5. Online Diary Quality by Income Group (n=586)



4.4.4. Logistic Regression Results

Logistic regression models were specified in order to validate the above findings and look further into the relationship between online diary quality and several household demographic indicators.

Figure 4.4.4. Logistic Regression Model of Demographic Indicators on Online Diary Quality (n=586)

	Model 1		*Model 2*		*Model 3*	
	Coefficient	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Intercept	22.10631	2180.6	3.077032	2.152687	2.795416	2.193353
Ref. age 25-34	-14.39232	2180.6	0.5989547	0.921237	0.8330859	0.9401892
Ref. age 35-49	-14.5641	2180.6	0.4635	0.864421	0.6204074	0.8729944
Ref. age 50-64	-15.2145	2180.6	-0.5279871	0.825828	-0.3795609	0.8405639
Ref. age 65+	-14.80172	2180.6	0.3488064	0.901216	0.7697971	0.9307495
CU size (1)	-0.3235674	0.4612	0.0999509	0.389914	0.2094546	0.4092014
CU size (3+)	-0.0048491	0.5197	-0.306683	0.438949	-0.2045507	0.447401
Log (pre-tax income)	-0.4314573	0.2486	-0.0902349	0.174402	-0.0487599	0.1796921
Reference person Hispanic	-0.1248928	0.6653	-0.2011475	0.533757	0.0297232	0.5673115
Reference person White	*1.057886	0.4328	*1.00369	0.381633	**1.047732	0.403247
Urban Area	-0.0893331	0.5451	-0.2359129	0.462711	-0.3012129	0.4812548
Homeowner CU	-0.9352134	0.6059	-0.1122071	0.425752	-0.1632558	0.4476053
Highest: Some college	0.1664486	0.4404	0.3821669	0.384519	0.3069787	0.4000238
Highest: HS grad or less	**-1.589319	0.4957	**-1.32603	0.38786	***-1.447796	0.4035579
Log (ZTOTAL)	0.1079706	0.1339				
Reference person male	-0.1137443	0.3241	-0.1137443	0.324096	-0.3195201	0.338798
Converted Refusal					***-2.252228	0.4600629
Model Pseudo R^2	0.086		0.086		0.1465	

*p<0.05, **p<0.01, ***p<0.0001

The logistic regression results in Figure 4.4.4 above helped solidify the previous findings that the race of the reference person, and the education level of the reference person were both significantly associated with the FR reported online diary quality. Across all three models the binary indicator for a reference person identifying their race as “White” was significantly more likely to be associated with an FR rating the diary as high quality. Conversely, the binary indicator for a reference person having an education level of “HS Grad or Less” was significantly more likely to be associated with a low quality online diary across all three regression models.

An interesting new finding in the third logistic regression model came from the inclusion of an indicator for converted refusals. Not only did this variable show to have a significant association with low quality online diaries, but it also improved the overall fit of the model. This may be an indicator that the FR is perceiving quality based on the respondent's overall cooperativeness and not necessarily objective quality measures. For all the models, the variables for age, CU size, income, Hispanic origin, housing tenure, total expenditures, and gender were not significantly associated with FR reported diary quality.

Recommendations on Targeting Online Mode

While several variables known to FRs prior to diary placement showed to have a significant association with online diary quality as rated by FRs, there are some potential roadblocks in using this information for targeting households for online diary placement. The number of reference persons in online diary cases identifying their race as Asian was so low that it may be unwise to use this result at scale for determining which households should be receiving an online diary, especially considering that there was no significant association with diary quality for this group when only looking at LSF paper diaries or all LSF cases. On the other hand, the significant relationship between education level and online diary quality could be promising as an online household targeting mechanism. It is possible that only offering online diaries to households where the highest education is greater than "High School Graduate" could result in a higher percentage of high quality online diaries. Given the results of the third logistic regression model in Figure 4.4.4, it may also be useful for bolstering online diary quality to only offer paper diaries to converted refusal cases.

As noted earlier in this section, the exact heuristics by which FRs assess and grade diary quality is not completely clear, and their ratings should not be taken as the final word on diary quality. This is exemplified in the section on allocation rates, which shows that LSF online diary entries were allocated at a higher rate than LSF paper diaries. These results would imply that LSF paper diaries were on average, of higher quality than online diaries, which directly opposes the results from comparing data quality assessments as shown in Figure 4.4.2.

5 RECOMMENDATIONS FOR IMPLEMENTATION AND LESSONS LEARNED

Based on analysis of the processed data from the LSF, we recommend officially implementing online diaries into production, while continuing to improve the protocols, FR training, and the online instrument to address the quality issues that have been identified. We recognize that the option to have an online diary during the pandemic has been essential for providing a no-touch mode. We also recognize that given the uncertain environment related to the pandemic, continuing to use the online mode is critical.

Several issues regarding protocols, FR training, and instrument design need to be addressed in order to provide higher response rates and better data quality in the online diaries. These include:

- Explore performance-based incentives or other methods to motivate online diary respondents to provide better quality data that would not require removal by minimal expenditure edit.
- Provide better help desk access and training for help desk staff to help with login issues.
- Make sure as many FRs as possible login with the respondent during the pickup visit.
- Provide a way for FRs to edit data reported in the online diary data, similar to how they would edit a paper diary.
- Allow FRs to disclose that they can view the respondent's entries to facilitate a conversation of encouraging higher data quality.
- Add features to the online diary instrument such as hover-over instructions and examples for "other" category in the online diary. Consider adding categories to split the "other" into more easily understood categories.
- Do not closeout the instrument at pickup to allow FRs more time to enter expenses from receipts.
- Investigate online placement procedures in more detail to see whether it makes sense to exclude online placement for certain categories such as those with an education level of high school or less as this group was rated by FRs as having lower data quality in the LSF.

Since the online diary has been in use in the Diary Survey as a contingency during the pandemic, Census has already made some progress in implementing recommendations from our preliminary report including changing the protocol for recall and working towards developing explicit placement procedures for paper diaries for online eligible respondents. Efforts are underway at BLS to test new features and designs to improve data quality under an Interagency Agreement with Census and we plan to investigate other features in the future.

APPENDIX A: DIARY SOURCED UCC DIFFERENCES BY GROUP

We examined the top 10 entries by cost for both the LSF and Production groups within the "Other" section (sub-setting to those sourced from the diary). The UCC codes from these CUs were identified, resulting in (after accounting for duplicates across groups and after combining similar UCCs) 14 UCC categories that were tallied by group. The share of all "Other" section entries containing these 'high-dollar' UCCs and the average amount reported were compared across groups to get a sense whether certain UCCs were driving the differences in reporting among Production and LSF groups (Figure A1). Among the most commonly reported high-dollar UCCs in Figure A1, we found few with large percent differences between Production and LSF.

Figure A1. High dollar "Other" section UCCs by group: Share reporting, mean amounts, and percent difference of means

Production				*LSF*			*Differences*			
UCC Label	*Freq*	*%*	*Mean value*	*Freq*	*%*	*Mean value*	*Share reports*	*Mean $ reported*	*% Diff.*	*Description*
Cosmetic	2163	7.78	$12	575	7.25	$17	0.53	-$5	34.5%	Cosmetics, perfume, bath preparations
Medical	1427	5.14	$14	348	4.39	$17	0.75	-$3	19.4%	Topicals and dressings Nonprescription vitamins
Toys	1079	3.88	$18	338	4.26	$23	-0.38	-$5	24.4%	Toys, games, arts and crafts, and tricycles
Home goods	1041	3.75	$38	253	3.19	$24	0.56	$14	45.2%	Telephones & accessories Clocks and other household decorative items
Lawn costs	495	1.78	$31	133	1.68	$50	0.1	-$19	46.9%	Lawn and garden equip -ment Lawn and garden supplies
Auto repair	341	1.23	$31	96	1.21	$31	0.02	$0	0.0%	Misc. auto repair,servicing
Sports	198	0.71	$67	49	0.62	$60	0.09	$7	11.0%	Fees for participant sports
Bath linens	191	0.69	$12	50	0.63	$18	0.06	-$6	40.0%	Bathroom linens
Pet medical	153	0.55	$179	31	0.39	$237	0.16	-$58	27.9%	Vet services
Hunting	107	0.39	$82	33	0.42	$55	-0.03	$27	39.4%	Hunting and fishing equip -ment
Infants' goods	62	0.22	$44	14	0.18	$34	0.04	$10	25.6%	Infants' equipment
Luggage	53	0.19	$85	9	0.11	$43	0.08	$42	65.6%	Luggage
Misc. fees	10	0.04	$289	2	0.03	$26	0.01	$263	167%	Miscellaneous fees
Repairs	7	0.03	$178	4	0.05	$151	-0.02	$27	16.4%	Repair of miscellaneous household equipment and furnishings
All Other UCCs	20,459	73.6	$12	5,994	75.6	$11	-1.96	$1	8.7%	