PROOF OF CONCEPT TEST FOR THE CONSUMER EXPENDITURE SURVEY

RESULTS ON INTERVIEW AND DIARY EXPENDITURE REPORTS AND DATA QUALITY

VERSION DATE: 4/6/2017

EDITED FOR PUBLIC RELEASE: 5/3/2017

Safia Abdirizak
Angi Lee
Brett McBride

U.S. Bureau of Labor Statistics
Division of Consumer Expenditure Survey
Branch of Research and Program Development

Table of Contents

I. Executive Summary 3
II. Report Scope 6
III. Overview 6
IV. Records and Recall Interviews Data Quality 7
a. Overall Expenditure Comparisons 7
b. Interview Expenditure Totals Controlling for CU Characteristics 7
c. Section-Level Expenditure Comparisons 9
d. Section-Level Counts of Expenditures 13
e. Response Quality – Data Editing and Missing Data Rates 14
f. Impact of Record Use on Records Interview Expenditures 15
g. Response Quality – Extent of Rounded Responses in Records Interview 18
V. Diary Data Quality 20
a. Overall and Section-Level Expenditure Comparisons 20
b. Number of Entries Controlling for CU Characteristics 21
c. Within-Household Reporting Characteristics 23
VI. Conclusion 28
APPENDIX 29

I. Executive Summary

This report focuses on the quality of data collected in the recall and records interviews and the personal diaries of the Proof of Concept (POC) test[1]. Data quality was measured by the extent of expenditures captured, the extent to which those expenditures were free from incomplete data, and the precision of the expenditure amounts provided, all of which were assessed through comparisons with data collected through the current Consumer Expenditure (CE) survey, referred to as the Restricted Production (RP) sample. Additional research explored the extent to which data quality may have benefited from respondents' use of records.

The report's main findings, summarized below, support a recommendation to move forward with changes to the survey as proposed in the Gemini redesign plan. Test findings suggest the redesigned survey would accomplish its goal of a verifiable reduction in measurement error. The findings met or exceeded all a priori thresholds for considering the POC test to be successful. The test led to higher overall interview expenditure totals though this was not significant when controlling for other factors. Although the test had diary expenditure totals that were not significantly higher, the number of personal diary entries was significantly higher when controlling for other factors. The interview and diary were also associated with improvements in interview data quality. This and prior reports do point to the need to pursue increased response rates as part of a redesigned survey, to further encourage record use to improve expenditure quality, and to continue working toward an improved version of the online diary.

Records and Recall Interview Data Quality

- Overall Expenditures Totals. The POC sample had significantly higher expenditure totals reported in the interview, with the average being $10,780 compared to $9,329 for the RP sample, and had median amounts that were 11 percent higher (Table 1). The average income levels of Consumer Units (CUs) – $71,526 for POC CUs and $66,666 for RP CUs – were not found to be significantly different, and missing data for income sources among RP CUs may have deflated the overall income average for that sample (Table 5).

- Expenditure Amounts Controlling for CU and Test Characteristics. A regression found that CUs participating in the POC test were not associated with significantly higher reported expenditure amounts, after controlling for other factors such as income and number of CU members (Table 2).

[1] The CE survey previously used 'individual diaries' to refer to this component of the redesign, but has transitioned toward use of the more commonly-used 'personal diaries.'

CUs that used records were associated with significantly higher expenditure amounts, but this finding did not reach significance when included in an interaction with the POC test condition.

- Count of Expenditures Reported. In the recall interview sections, POC CUs tended to report more expenditures than their RP counterparts (Table 10). The record sections did have some sections (e.g., insurance) where RP CUs reported significantly more expenditures – although few sections involved statistically significant differences (Table 11).

- Section-Level Expenditure Reports. Two of the 5 recall sections involved significantly higher expenditure totals in the POC sample than in the RP sample (Table 3). Among record sections, as with the expenditure counts, the Rented Housing section had higher expenditures totals for POC CUs, while the Health Insurance section had a higher expenditure average for RP CUs (Table 4).

- Measures of Response Quality. POC respondents provided fewer missing values for expenditure variables than did RP respondents (Tables 13 and 14). In addition, the extent of values needing editing during CE data processing was lower for POC than for RP CUs (Table 12). Overall, there was significantly less rounding in the POC sample across questions in the records interview. In the RP sample, CUs rounded an average of 34 percent of the expenditures examined, whereas POC CUs rounded an average of 22 percent of expenditures (Table 20).

- Effect of Record Use. POC sample CUs with section expenditures who were reported to use records in the record sections did not have significantly larger expenditure averages for many of the sections (Table 18). For the RP sample, although record use was less prevalent, record use was reliably associated with higher section averages compared to no record use (Table 19).

Diary Data Quality

- Overall Expenditure Totals. Despite the larger number of diary entries provided by POC CUs, there was not a significant difference in the total expenditures reported by this sample. POC sample CUs had an average of $1,096 in weekly diary expenditures recorded in their diaries compared to an average of $1,037 for RP sample CUs, and had a median amount that was 9 percent lower than the RP median (Table 22). The lower average income levels of POC diary CUs may have affected this finding (Table 23).

- Section-Level Expenditure Totals. There were no significant differences found at the section level; the 'Food Away from Home' section had higher median values for POC CUs than RP CUs, but POC CUs had lower or equal median values in the other three sections (Table 24).

- Count of Entries Controlling for CU and Test Characteristics. A regression found that CUs participating in the POC test were associated with a 31 percent greater number of expenditure reports than those in the RP sample, after controlling for factors such as income, CU size, and education. Of interest, CUs in which all members were placed with an online diary were associated with 48 percent fewer expenditure reports than other CUs and POC CUs with multiple members had a significant association with increased expenditure reports (Table 25).

- Patterns of Within-Household Reporting. In POC CUs, certain diarists tended to provide a large portion of the expenditures reported by multi-diary CUs (Table 27). These diarists tended to be 35-to-64 years old (this age group made up 68 percent those who supplied 51 or more entries). Diarists who were present when Field Representatives (FRs) instructed CU members on the diary task were associated with twice the number of expenditure reports than those who were not (Table 29). Although diarists age 20 or younger were disproportionately likely to not provide a complete diary, expenditure descriptions suggested these diarists reported small expenditures that would not have been captured through a one-diary-per-CU design (Table 28).

II. Report Scope

The results and recommendations from the Proof of Concept (POC) test were detailed in separate reports:

Report 1: written jointly by Census and BLS and focused on training, field procedures and the Field Representative (FR) Debriefing summary. The report included recommendations for training, protocols, and design as well as preliminary response rate estimates (Bagley et al., 2016).

Report 2: written by BLS, focuses on analyzing POC data that have been processed by the BLS Initial Edit System. The report included detailed analysis of BLS defined response rates, contact attempts, diary use, and analysis of FR and respondent debriefing questions (Erhard et al., 2016).

Report 3 (this report): written by BLS and focusing on analyzing POC data that had been processed by the BLS Edit and Estimation System. This report includes detailed analysis of total expenditures and data quality in the POC data.

III. Overview

As part of the Consumer Expenditure (CE) Survey's redesign plan, CE tested components of a modernized version of the survey. These components included one sample for both interview and diary surveys, the provision of incentives and an emphasis on record use. The POC interviews were composed of most of the sections from the current survey, but divided into a recall interview and a records-based interview. Similarly, the diary survey was modernized to incorporate the choice of using online diaries and provided personal diaries for eligible members of the household. More information on the test can be found in the overview section of the prior report.

The analysis carried out for this report was of the sample of 'completed cases' – the 520 households completing both of the interviews and the diary[2]. Expenditure totals, CU characteristics and data quality for this sample was compared against production data collected in the same time period as the POC test (2015 quarter 3) and in only the same counties sampled for the POC test, and for the diary involved the same 1-week reporting period. This 'Restricted Production' (RP) sample was composed of 1,483 interview cases (across all waves) and 315 diary cases. In order to make equivalent comparisons, this report only examined the RP interview sections that were administered as part of the POC test[3].

[2] Diary completion was defined by having at least one expenditure recorded or provided via diary recall, or indication of a week without expenditures. In addition, the total number of expenditures reported via the diaries or diary recall needed to meet productions' processing thresholds for the minimal expenditure edit.

[3] The POC test did not involve asking questions from interview sections 5, 7, 9, 12, 16 and 20. More information about data processing for the POC test can be found in the Appendix.

IV. Records and Recall Interviews Data Quality

a. Overall Expenditure Comparisons

The POC test involved two FR visits in which FRs collected the expenditures described below. An initial visit involved FRs collecting expenditures incurred over the prior 3 months via recall. Following collection of the weekly diaries, the second interview visit collected expenditures from sections where records would be available to assist with reporting accuracy. We computed the amount of expenditures reported by the RP sample for the same expenditure sections collected in the first and second POC visit, to compare against expenditures reported in the POC visits. In the analysis plan for the POC, the team set guidelines to evaluate success of the POC test. For this analysis, it held that the POC's median total diary expenditure must be no lower than 10 percent of the median for RP. The table below indicates that POC sample CUs reported significantly higher average overall expenditures than RP sample CUs[4].

Table 1. Comparison of CU expenditure totals by sample

	N	Mean	Median
POC	520	$10,780	$7,457
RP	1480	$9,329	$6,697

The CUs in the POC sample reported an average of $10,780 in expenditures compared to the $9,329 average reported by CUs in the RP sample, and also had a higher median value than the RP sample ($7,457 compared to $6,697). The evaluation guideline was met, with the POC sample having a median expenditure total 11 percent higher than that of the RP sample.

b. Interview Expenditure Totals Controlling for CU Characteristics

We regressed the log of expenditure totals on the sample condition and associated CU characteristic variables. We included the following variables in the model:

- presence in a multi-member CU,
- income and lack of income ('all valid blanks'),
- age,
- record use,
- education level of the member with the highest education level,
- CU tenure, and
- race of respondent

[4] Wilcoxon-Mann-Whitney; Z= 2.46, Pr>Z 0.014. See Appendix for more information on statistical methods used. 3 CUs were missing from the RP sample as they were transmitted with no more follow-up possible (Outcome=203).

We produced an additional model that explored a two-way interaction between CUs in the POC sample and CUs with some record use. The final regression models with only the significant variables included are presented below (Table 2).

Table 2. Model of log (interview expenditure totals) by sample condition and associated covariates (n=1,994[5])

	Model 1		Model 2	
	Coefficient	SE	Coefficient	SE
Intercept	8.038	(0.071)	8.046	(0.071)
Income	0.319***	(0.016)	0.318***	(0.016)
Multi-member CU	0.315***	(0.033)	0.314***	(0.033)
Homeowner	0.167***	(0.031)	0.168***	(0.032)
POC CU	0.028	(0.035)	-0.077	(0.083)
Education Level	0.152***	(0.014)	0.152***	(0.014)
Race of Respondent	-0.141***	(0.036)	-0.144***	(0.036)
At Least Some Record Use	0.156***	(0.031)	0.140***	(0.033)
All Valid Blanks	-0.701***	(0.066)	-0.702***	(0.066)
POC x Record Use	--	--	0.126	(0.091)
Model R^2	0.461		0.462	

*p<.05; **p<.01; ***p<0.001

Model 1 indicated a number of the variables to have the expected relationship with expenditures; higher income[6] and education levels, multi-member CUs and owned homes all were associated with greater expenditure totals. CUs that consulted records were associated with 16 percent higher expenditure totals compared to CUs that used records less than 10 percent of the time. When controlling for those factors and the respondents' race, being in a POC CU was associated with higher expenditures, although this was not significant. As participation in the POC test was closely related to reported record use[7], Model 2 included a two-way interaction between POC CUs and CUs with record use. Although this interaction was not significant, it suggests a positive association between record use within POC CUs and higher expenditure reports.

[5] Nine RP CUs were excluded from the regression - three CUs did not report expenditures (Outcome=203) and six CUs had missing values for the Record use variable. See the Appendix for variable frequencies.
[6] Since $0 values for the income variable captured some CUs that provided 'Don't Know' or 'Refuse' responses, we included a separate variable 'All valid blanks,' to designate CUs with legitimate reports of zero income across all income sources – as expected, these CUs were associated with a 70 percent lower expenditure total.
[7] 88 percent of POC CUs reportedly had record use; whereas 43 percent of RP CUs were reported to use records greater than 10 percent of the time. Note, these variables were not found to be collinear.

c. Section-Level Expenditure Comparisons

Further analysis was carried out to compare the expenditure totals reported in each interview section. The tables below show expenditure totals aggregated to the section level for the recall and records interviews[8]. More information on the methods of calculating expenditure totals by section can be found in the Appendix.

Table 3. Comparison of section expenditure totals: recall interview sections

	POC (N=479)		RP (N=1,313)		
	Mean	Median	Mean	Median	Wilcoxon Z
6/Appliances	$399	$90	$267	$1	5.18***
8/Furnishings	$380	$43	$282	$0	5.94***
17/Subscriptions	$234	$72	$222	$60	1.76
18/Trips	$653	$0	$707	$0	1.17
19/Miscellaneous	$696	$180	$950	$330	-5.77***

*p<.05; **p<.01; ***p<0.001

Table 4. Comparison of section expenditure totals: records interview sections

	POC (N=520)		RP (N=1,480)		
	Mean	Median	Mean	Median	Wilcoxon Z
2/Rented Housing	$1,083	$0	$923	$0	2.10*
3/Owned Housing	$1,664	$656	$1,783	$812	-1.11
4/Utilities	$1,258	$1,200	$1,290	$1,242	-0.62
10/Rented and Leased Vehicles	$116	$0	$156	$0	-0.65
11/Owned Vehicles	$1,284	$0	$1,404	$0	0.43
13/Non-Health Insur.	$454	$239	$518	$279	-1.78
14/Health Insur.	$587	$260	$772	$499	-6.57***
15/Medical	$336	$30	$334	$47	-1.31

*p<.05; **p<.01; ***p<0.001

These tables indicate that expenditure reports varied between the samples depending on the interview in which they were collected. Two of the 5 recall sections involved significantly higher expenditure totals in the POC sample than in the RP sample, with there being only 1 section (Section 19/Miscellaneous) in which the RP sample reported significantly higher expenditure totals[9]. Among the record sections, only

[8] Negative values could result from CUs reporting reimbursed amounts. CUs with no expenditures in each section were excluded (see the Appendix for more information). Record sections that did not involve 3-month expenditures (e.g., Section 1/Housing characteristics, Section 21/income, or Section 22/assets) were not included in the analysis.
[9] Significance based on Wilcoxon-Mann-Whitney tests, Z test statistics represented.

the Rented Housing section had a significantly higher expenditure total in the POC sample than in the RP sample, while Health Insurance had a higher expenditure total reported by RP sample CUs.

To determine whether overall expenditure differences may have been affected by the relative incomes among the POC and RP sample CUs, we compared total CU pre-tax income across samples (Table 5).

Table 5. CU income (FINCBTAX) by sample

	N	Average	Median
POC Income	520	$71,526	$45,404
RP Income	1,483	$66,666	$43,459

The results from the statistical significance test revealed that the income levels of the two samples were not significantly different[10]. This alleviates concerns that providing incentives might disproportionately skew the POC sample composition toward CUs with lower income levels. The results above may have been affected by how pre-tax income (FINCBTAX) is calculated[11]. Missing values for income sources are treated as zero values in processing, which can result in this variable indicating smaller values for income (Table 5) if more missing values are present. Table 6 below further analyzes the income source variables, displaying the percent of CUs with missing values by sample.

Table 6. Extent of CU missing values for income sources by sample[12]

	POC		RP	
	N (CUs)	% missing values	N (CUs)	% missing values
Wages and salaries*	406	7.1%	1128	10.7%
Social Security or Railroad Retirement income*	134	3.7%	406	9.9%
Interest and dividends income	111	10.8%	279	17.2%
Retirement, survivor, or disability pensions	80	2.5%	188	12.8%
Self-employment income*	59	30.5%	186	41.4%
Food stamps	53	3.8%	137	5.8%
Other income source (VA payments, unemployment compensation, child support, or alimony)	32	3.1%	80	6.3%
Supplemental Security income*	19	0%	74	4.1%
Net rental income	26	11.5%	58	6.9%
Royalty income /income from estates and trusts	9	11.1%	26	15.4%
Care of foster children, cash scholarships, and fellowships	13	0%	16	0%
Public assistance or welfare	4	0%	16	25.0%

*For these sources, the percent indicates that all of the members of the CU reported a missing value (see Table 7)

[10] Wilcoxon-Mann-Whitney; Z=1.61, Pr>Z 0.1086
[11] Imputed income measures were not available for the POC test.
[12] For each income source, we excluded CUs that reported no income source (i.e., excluded valid blanks)

The percent of missing values in income variables was higher or the same for RP CUs than for POC CUs for every item except "Net rental income" (Table 6). Since some income sources (those designated by an asterisk in the table above) were derived from responses provided at the member level, we examined missing values for sources at that level (Table 7). Income sources were categorized into: "none missing" (e.g., all members of the household reported an income value), "1+ missing" (at least one member of the household reported a missing value), and "all missing."

Table 7. Extent of CU missingness for member-level income sources by sample

	POC				RP			
	N	None missing	1+ missing	All missing	N	None missing	1+ missing	All missing
Wages and salaries	406	88.4%	4.4%	7.1%	1128	84.2%	5.1%	10.7%
Self-employment income	59	66.1%	3.4%	30.5%	186	55.9%	2.7%	41.4%
Supplemental security income	19	100.0%	0%	0%	74	95.9%	0%	4.1%
Social Security or Railroad Retirement income	134	95.5%	0.7%	3.7%	406	88.9%	1.2%	9.9%

The percent with partial or full within-CU missing values was lower for POC than for RP. Similarly, there were higher percentages in the "none missing" column for POC than for RP (Table 7). The large number of missing values among RP CUs in Tables 6 and 7 could imply that incomplete income reporting may have affected the income estimates, such that the true mean value of RP CUs' income could be higher than the results shown in Table 5.

Table 8. Proportion reporting income sources

Income Source	POC (N=520)	RP (N=1483)	Diff. (POC-RP)
Wages and salaries	78.1%	76.1%	2.0%
Social Security or Railroad Retirement income	25.8%	27.4%	-1.6%
Interest and dividends income	21.3%	18.8%	2.5%
Retirement, survivor, or disability pensions	15.4%	12.7%	2.7%
Self-employment income	11.3%	12.5%	-1.2%
Food stamps	10.2%	9.2%	1.0%
Other income source (VA payments, unemployment compensation, child support, or alimony)	6.2%	5.4%	0.8%
Supplemental Security income	3.7%	5.0%	-1.3%
Net rental income	5.0%	3.9%	1.1%
Royalty income or income from estates and trusts	1.7%	1.8%	0.0%
Care of foster children, cash scholarships, and fellowships	2.5%	1.1%	1.4%
Public assistance or welfare	0.8%	1.1%	-0.3%

Table 8 displays the proportion of CUs reporting various income sources. There were a higher proportion of RP CUs reporting income sources in Social Security, Self-employment, Supplemental Security income, Royalty income, and Public assistance income than POC CUs.

We further tabulated the mean and median values of the income sources reported to provide a more detailed look at income reporting in the POC test (Table 9).

Table 9. CU income source reports by sample*

	POC			RP			
Income Source	N (CUs)	Mean	Median	N (CUs)	Mean	Median	% diff. median (POC-RP)
Wages and salaries**	333	$78,769	$56,000	827	$80,068	$59,000	-1.6%
Social Security or Railroad Retirement income**	90	$1,610	$1,500	226	$2,217	$1,713	-37.7%
Interest and dividends income	75	$4,211	$37	159	$4,538	$180	-7.8%
Retirement, survivor, or disability pensions	72	$14,944	$5,264	148	$17,224	$6,000	-15.3%
Self-employment income**	34	$32,487	$5,508	77	$32,868	$10,000	-1.2%
Food stamps	51	$306	$224	129	$239	$192	21.9%
Other income source (VA payments, unemployment compensation, child support, or alimony)	30	$10,594	$3,714	72	$6,356	$3,020	40.0%
Supplemental Security income**	18	$3,617	$967	63	$4,246	$1,247	-17.4%
Net rental income	22	$6,063	$1,450	50	$5,681	$3,753	6.3%
Royalty income or income from estates and trusts	7	$5,585	$2,000	20	$17,554	$1,650	-214.3%
Care of foster children, cash scholarships, and fellowships	12	$8,789	$4,700	15	$13,593	$5,000	-54.7%
Public assistance or welfare	4	$4,073	$3,900	10	$2,768	$1,394	32.0%

* Excludes CUs reporting 'Don't Know'/'Refuse' to amount (i.e. those asked for bracketed income range), but includes those with a $0 response (e.g. one member reporting no income from source but another reporting amount).
** Aggregated to the CU-level

There were higher median totals reported for Social Security, Interest, and Retirement for RP CUs than POC CUs – those are income sources commonly associated with CUs on fixed income. Note there are small sample sizes for some income sources at the bottom of Table 9, as they may have been further reduced by a large number of CUs not reporting or knowing an exact amount.

d. Section-Level Counts of Expenditures

We computed the counts of expenditures reported in the sections of the two POC interviews (Tables 10 and 11). This involved removal of some edited or duplicated expenditures as described in the Appendix. The findings from these comparisons were largely the same as those for expenditure totals.

Table 10. Comparison of section entry totals: recall interview sections

	Overall (POC) (N=520)		Overall (RP) (N=1,483)		
	Mean	Median	Mean	Median	Wilcoxon Z
6/Appliances	1.56	1	1.00	0	6.11***
8/Furniture	1.32	1	0.85	0	6.80***
17/Subscriptions	2.51	1	2.15	1	3.02**
18/Trips	1.95	0	1.84	0	1.70
19/Miscellaneous	2.01	1	2.56	2	-3.49***

*p<.05; **p<.01; ***p<0.001

The recall interview sections involved the same finding of generally higher entry totals for the POC CUs than for the RP CUs. Unlike with expenditure amount comparisons (Table 3), entries were significantly higher in the POC sample than in the RP sample for Section 17/Subscriptions. As with expenditure comparisons, however, RP CUs reported significantly more entries in Section 19/Miscellaneous.

Table 11. Comparison of section entry totals: record interview sections

	Overall (POC) (N=520)		Overall (RP) (N=1,483)		
	Mean	Median	Mean	Median	Wilcoxon Z
2/Rented Housing	0.42	0	0.37	0	1.91
3/Owned Housing	1.77	2	1.86	2	-0.96
4/Utilities	4.01	4	3.92	4	0.89
10/Rented and Leased Vehicles	0.11	0	0.13	0	-0.58
11/Owned Vehicles	0.43	0	0.43	0	0.38
13/Non-Health Insur.	1.13	1	1.26	1	-2.00*
14/Health Insur.	0.91	1	1.22	1	-6.57***
15/Medical	1.58	1	1.48	1	-1.02

*p<.05; **p<.01; ***p<0.001

Examining the records interview findings, differences in average entries tended to be small; in some sections counts being higher for RP CUs, albeit non-significantly. Unlike the record section expenditure comparisons, Section 2/Rented Housing did not have significantly higher reports in POC (though this

approached significance, p=0.057) while Section 13/Insurance did have significantly higher counts in the RP sample than in the POC sample (in addition to Section 14/Heath Insurance).

e. Response Quality – Data Editing and Missing Data Rates

As a part of data quality measurement, we compared the proportions of data edited, allocated, and imputed during the production processes for both POC and RP samples (Table 12). The evaluation guidelines indicated that the POC editing rates must be no higher than 10 percent of the RP rates. Our findings showed that all of POC's editing rates were lower than RP's rates.

Table 12. Percent of cost values needing data editing by sample

	N (UCCs)	% edited	% allocated	% imputed	% combined
POC	26,751	15.0%	6.1%	8.4%	0.1%
RP	77,941	21.9%	7.4%	13.2%	0.2%

As Table 12 shows, fewer cost amounts needed editing, allocating and imputing for POC sample responses than for RP sample responses. We also examined the number of non-substantive values in both POC and in comparable RP sections. There are two respondent-provided invalid blanks collected, "Don't Knows" (DKs) and "Refusals" (RFs). The number of DK and RF expenditure items throughout the interview were summed for each CU. The means are reported in Tables 13 and 14 and the distribution of the missing values per CUs are shown in Tables 15 and 16.

Table 13. Number of missing values: Don't Knows (DKs)

	Mean
POC (N=520)	0.59
RP (N=1483)	0.92

Table 14. Number of missing values: Refusals (RFs)

	Mean
POC (N=520)	0.04
RP (N=1483)	0.13

There were an average of 0.6 DKs per CU in the POC sample and a significantly higher number of DKs, 0.9, in RP[13]. The difference in the average number of RFs between POC CUs and RP CUs was 0.09, also

[13] Wilcoxon-Mann-Whitney; Z=-4.04, Pr>Z <0.0001.

a statistically significant difference[14]. To provide more insight into missing values, we categorized the number of DK and RF responses (Tables 15 and 16).

Table 15. Distribution of number of Don't Knows (DKs) by sample

	POC (N=520)		RP (N=1,483)	
# of DKs	Number of CUs	Percent of CUs	Number of CUs	Percent of CUs
0	408	78.5%	1,029	69.4%
1	56	10.8%	204	13.8%
2-5	46	8.9%	189	12.7%
6-10	7	1.4%	42	2.8%
11-30	3	0.6%	19	1.3%
Total	520	100%	1,483	100%

Table 16. Distribution of number of Refusals (RFs) by sample

	POC (N=520)		RP (N=1,483)	
# of RFs	Number of CUs	Percent of CUs	Number of CUs	Percent of CUs
0	515	99.0%	1,430	96.4%
1	3	0.6%	24	1.6%
2-5	0	0%	21	1.4%
6-10	1	0.2%	3	0.2%
11-30	1	0.2%	5	0.3%
Total	520	100%	1,483	100%

Among CUs who provided values for expenditure related questions, those in the POC sample were more likely not to have any DK responses – 79 percent having none – compared to the comparison group – 69 percent. Similarly, the percent of CUs with no refusals for these questions in the POC sample was 99 percent, compared to 96 percent in RP (Table 16). When looking at the DK response distributions, the percentage of RP sample CUs with multiple responses was higher than in the POC sample, indicating the lower extent of missing values that needed editing in POC.

f. Impact of Record Use on Records Interview Expenditures

The POC records interview encouraged CUs to utilize records when reporting expenditure values. POC was designed in a way that FRs indicated whether a CU used records upon the completion of each section[15]. Table 17 shows the percentage frequency of record use among CUs with at least one expenditure in the records interview.

[14] Wilcoxon-Mann-Whitney; Z=-3.06, Pr>Z <0.0023.
[15] The current CE interview only asks about record use at the end of the interview and only for certain sections/section combinations, limiting the record use comparisons possible between POC and RP samples.

Table 17. Percentage frequency of record use for CUs with at least 1 expenditure

POC	Percent	Total
3/Owned Housing	65%	312
4/Utilities	78%	512
10/Rented Vehicles	50%	52
11/Owned Vehicles	64%	174
13/Insurance	73%	344
14/Health Insurance	67%	302
15/Medical	54%	280

RP	Percent	Total
3/Owned Housing	16%	936
4/Utilities	36%	1,456
13/14 Insurance	18%	1,480
15/Medical	23%	877

Table 18 shows that, on average, POC CUs that consulted records were more likely to have higher section totals than POC CUs that did not[16].

Table 18. POC sample: comparison of expenditure totals by record use (among CUs with 1+ section expenditure)

	With Section Record Use (POC)			Without Section Record Use (POC)			
	N	Mean	Median	N	Mean	Median	Wilcoxon Z
3/Owned Housing	203	$2,835	$2,121	109	$2,660	$1,949	-0.64
4/Utilities	401	$1,292	$1,220	111	$1,226	$1,200	-0.80
10/Rented and Leased Vehicles	26	$959	$905	26	$1,363	$422	-1.78
11/Owned Vehicles	111	$4,027	$392	63	$3,505	$306	-1.20
13/Non-Health Insur.	252	$736	$450	92	$552	$389	-2.31*
14/Health Insur.	201	$1,047	$676	101	$936	$585	-1.62
15/Medical	151	$736	$276	129	$494	$170	-4.11***

*p<.05; **p<.01; ***p<0.001

The average expenditure totals for the non-health insurance[17] and medical[18] sections were significantly higher for CUs that used records. However, the use of records was not associated with expenditure totals

[16] We limited analysis to CUs having at least one expenditure in a section (unlike how it was reported in report 2), as other CUs were unlikely to have used records. Record use was not collected for Section 2.
[17] Wilcoxon-Mann-Whitney; Z= -2.31 Pr>Z 0.0213
[18] Wilcoxon-Mann-Whitney; Z= -4.11 Pr>Z 0.0001

that were significantly different in the other sections. Section 10, 'Rented and Leased Vehicles,' did not have a higher average expenditure for POC CUs that consulted records. However, this may have been caused by an extreme value among POC CUs without section record use. A separate analysis, examined these differences for the number of entries (Table 18b).

Table 18b. POC sample: comparison of entries by record use (among CUs with 1+ section expenditure)

	With Section Record Use (POC)			Without Section Record Use (POC)			
	N	Mean	Median	N	Mean	Median	Wilcoxon Z
3/Owned Housing	204	2.97	3	110	2.86	3	-0.94
4/Utilities	400	4.26	4	111	3.41	3	-4.23***
10/Rented and Leased Vehicles	26	1.19	1	26	1.04	1	-1.40
11/Owned Vehicles	114	1.25	1	65	1.26	1	0.23
13/Non-Health Insur.	252	1.83	2	92	1.36	1	-4.22***
14/Health Insurance	201	1.70	1	101	1.30	1	-3.35***
15/Medical	150	3.61	3	128	2.19	2	-4.82***

*p<.05; **p<.01; ***p<0.001

In terms of entries, all sections (except Owned Vehicles) had higher entry averages when CUs used records to report expenditures. Differences were significant for Sections 4, 13, 14 and 15.

We also examined the impact of record use for the RP sample. Table 19 compares expenditure totals by section record use. The RP interview does not uniquely capture record use in as many sections (e.g., combining across certain sections), leading to the smaller number of sections examined below.

Table 19. RP sample: comparison of expenditure totals by record use (among CUs with 1+ section expenditure)

	With Section Record Use (RP)			Without Section Record Use (RP)			
	N	Mean	Median	N	Mean	Median	Wilcoxon Z
3/Owned Housing	151	$3,561	$2,561	785	$2,676	$1,955	2.80**
4/Utilities	519	$1,429	$1,358	937	$1,246	$1,188	5.27***
13&14/ Insurance	259	$1,916	$1,515	1221	$1,157	$789	6.22***
15/Medical	198	$985	$363	679	$442	$182	5.68***

*p<.05; **p<.01; ***p<0.001

The use of records was noted to be more common in the utilities and medical sections (in which 36 percent and 22 percent of RP CUs were reported to use records). When records were used, these sections involved significantly higher expenditure totals – average expenditure totals of $1,429 and $985, respectively. Across all sections, higher average expenditure totals were associated with record use. Table 19b shows the number of section entries for record sections that uniquely captured record use.

Table 19b. RP sample: comparison of entries by record use (among CUs with 1+ section expenditure)

	With Section Record Use (RP)			Without Section Record Use (RP)			
	N	Mean	Median	N	Mean	Median	Wilcoxon Z
3/Owned Housing	152	3.2	3	793	2.9	3	4.25***
4/Utilities	519	4.6	4	935	3.7	4	8.62***
13&14/ Insurance	228	1.9	2	865	1.6	1	4.81***
15/Medical	197	3.4	3	673	2.3	2	7.00***

*p<.05; **p<.01; ***p<0.001

As with Table 19, we found consistently higher entries associated with record use. It should be noted that, while we saw an association between record use and higher expenditure amounts and entries for most sections, this relationship may not be causal. In the absence of an experiment, we could also hypothesize that respondents who used records had different characteristics than those who did not, and that those characteristics were what lead to the reporting differences found in this analysis.

g. Response Quality – Extent of Rounded Responses in Records Interview

With respect to response quality, we measured the amount of rounded expenditures across sections in the records interview. Incidence of rounding has been used to evaluate data quality in survey data as demonstrated in a recent study on financial record checking conducted by RTI International, which suggests better quality data can be collected with little risk to respondent cooperation[19]. The study found evidence that the act of checking records was associated with lower levels of rounding on several measures. We defined a rounded expenditure as any value with 00, 25, 50 or 75 as the trailing digits[20]. To measure the extent of rounded responses in the records interview, we summed by CU the total count of

[19] Financial record checking in surveys: Do suggestive prompts improve data quality? Field Methods, 28(3), 247–265, Murphy, J. J., Rosen, J., Richards, A. K., Riley, S., Peytchev, A., & Lindblad, M. (2016)

[20] Responses of $0 were not treated as rounded values as they could represent missing values. Variables in the following sections were examined: health insurance, income, medical, non-health insurance, owned homes, rented and leased vehicles, and utilities.

values defined as rounded. The number of rounded values was then divided by the total number of reported expenditures by CU to derive a percentage of rounded expenditures (Table 20).

Table 20. CU average percent of rounded expenditures by sample

	N (CUs)	% Rounded	# Rounded	# Total
POC	510	22.1%	739	3,338
RP	1,432	33.8%	2,909	8,607

Table 20 indicates that for POC, the average CU rounded 22 percent of the expenditures that we observed. For RP, the average CU rounded 34 percent of their reported expenditures for observed variables. The evaluation guidelines indicated that rounding in POC must be 10 percent lower than in RP. The CU average percent of rounded expenditures surpassed the threshold, with 35 percent less rounding in POC CUs, a significant difference[21]. We further explored incidences of rounding in the records interview by question (Table 21).

Table 21. Comparison of CU average percent of rounded expenditures by question

	POC			RP		
	# Rounded	# Total	% Rounded	# Rounded	# Total	% Rounded
3/Owned Housing (MRTPMTX)	41	221	18.6%	263	597	44.1%
4/Utilities (UTLCHGX2)	118	946	12.5%	631	2,706	23.3%
10/Rented Vehicles (RENTEXPX)	12	26	46.2%	42	72	58.3%
13/Non-health Insurance (INSNEXXB)						
Life	17	151	11.3%	59	349	16.9%
Long Term Care	3	13	23.1%	8	35	22.9%
Automobile	47	214	22.0%	140	532	26.3%
Homeowners'	20	70	28.6%	54	158	34.2%
Renter's	1	29	3.4%	4	56	7.1%
14/Health Insurance (HHIRPMXB)	69	435	15.9%	287	1,121	25.6%
15/Medical (MEDPMTX)	204	821	24.8%	762	2,154	35.4%

Table 21 indicates lower levels of rounding in POC compared to the RP sample. Notably, sections where POC CUs had higher record use compared to the RP sample showed lower levels of rounding. For example, 78 percent of CUs in the POC sample referred to records in the utilities section compared to 36 percent of CUs in the RP sample (Table 17). In the table above, POC CUs answering a question about their utility expenses rounded 13 percent of responses, compared to 23 percent of responses being rounded among RP CUs.

[21] Wilcoxon-Mann-Whitney; Z=-8.16 Pr>Z <0.0001.

V. Diary Data Quality

a. Overall and Section-Level Expenditure Comparisons

Report 2 described how the one-week diaries provided to eligible household members as part of the POC test resulted in a higher number of overall entries, lower missing data rates and high rates of within-household completion compared to the equivalent RP sample. This section of the report expands upon those findings by presenting the expenditure totals reported – overall and by section, and identifying whether POC diary CUs had more entries after controlling for other factors. This section concludes by examining characteristics of reporting within CUs.

POC and RP sample CUs reported similar expenditure totals in the diary, when excluding CUs reporting all expenditures to the FR through recall or receipts – 'total recall' (Table 22). The evaluation guidelines indicated that the POC's median total diary expenditure must be no lower than 10 percent less than RP's. This was met for total expenditure amounts, as POC CUs had a median that was 9 percent less than RP's.

Table 22. CU total expenditure amounts by sample (excluding 'total recall' CUs)

	N	Mean	Median
POC	508	$1,095.55	$526.64
RP	259	$1,036.68	$579.09

POC sample CUs reported an average of $1,096 in weekly diary expenditures compared to an average of $1,037 for RP sample CUs, although POC CUs had a lower median amount. This difference was not statistically significant[22]. We also examined the totals when total recall CUs were included, with the same finding of non-significant differences between POC and RP expenditures (see the Appendix). As with the interview, we examined the income of the samples to see if that may have affected the comparisons above (Table 23).

Table 23. CU income (FINCBEFX) by sample[23]

	N	Mean	Median
POC Income	520	$67,154.46	$43,000.00
RP Diary Income	283	$74,057.31	$50,000.00

[22] Wilcoxon Mann-Whitney; Z=1.10, Pr>Z = 0.27. Note that Table 22 includes CUs with the 'no expense checkbox' checked ('legitimate blank diaries') but with ZTOTAL values as a result of data processing.
[23] The POC income amounts were slightly lower than those reported in Table 5. Diary income is drawn from a different variable, FINCBEFX, which is not 'prior to deductions for Medical insurance and Medicare.'

POC CUs had median incomes that were lower than that of RP CUs, though this difference was not significant[24]. Income characteristics were controlled for in the regression described in Section Vb.

In combination with report 2 findings of more diary entries for POC sample CUs, the lower median expenditure finding may also indicate that POC CUs reported more items with lower expenditure amounts. We further examined section expenditures in Table 24.

Table 24. CU section expenditure amounts by sample (excluding 'total recall' CUs)

	POC (N=508)		RP (N=259)		Wilcoxon Z
	Mean	Median	Mean	Median	
Food (Home)	$109.88	$76.22	$111.90	$82.76	0.37
Food (Away)	$81.03	$49.42	$75.08	$45.00	-1.85
Clothing	$43.23	$0.00	$51.32	$0.00	-1.65
Other	$861.40	$273.30	$798.38	$381.52	1.34

POC sample CUs reported an average of $110 in food for home consumption, similar to the $112 reported among their RP sample counterparts. For the 'Food Away from Home' section, POC households reported $81 on average compared to the $75 reported by RP households, but this difference was not significant (p=0.06). Although differences in section expenditures appeared large between the samples for the clothing and other sections, with a higher mean value for the RP sample for the clothing section and a higher mean value for the POC sample for the other section, the overall section distributions were not significantly different.

b. Number of Entries Controlling for CU Characteristics

Although report 2 found a larger number of diary entries for POC sample CUs, in this report we further controlled for other potentially explanatory factors. We regressed the log of entries for CUs that had one or more entry (reported via the diary or via recall/receipts) on the sample condition – POC or RP – and other associated variables. We initially included the following variables in the model:

- presence in a multi-member CU,
- all entries provided via recall/receipts ('total recall diaries'),
- categorized income (FINCBEFX),
- diary mode,
- education level of the member with the highest education level,
- CU tenure, and
- reference person race category.

[24] Wilcoxon Mann-Whitney; Z=1.69, Pr>Z = 0.09

We produced an additional model that explored two-way interactions. The regression for the final models with only the significant variables included is presented below (Table 25).

Table 25. Model of log(diary entries) by sample condition and associated covariates (n=803[25])

	Model 1		Model 2	
	Coefficient	SE	Coefficient	SE
Intercept	2.450***	(0.100)	2.632***	(0.116)
POC CU	0.313***	(0.068)	0.056	(0.113)
Multi-member CU	0.454***	(0.069)	0.241*	(0.108)
All entries via recall/receipts	-0.734***	(0.146)	-1.042***	(0.178)
Income category	0.100***	(0.026)	0.097***	(0.026)
Online diary CU	-0.478***	(0.085)	-0.494***	(0.084)
Education level	0.068*	(0.032)	0.066*	(0.032)
POC x Multi-member interaction	--	--	0.326*	(0.132)
POC x All entries via recall/receipts	--	--	0.854**	(0.302)
Model R^2	0.18		0.19	

*p<.05; **p<.01; ***p<0.001

Model 1 indicates that CUs participating in the POC test condition reported a larger number of diary entries than those in the RP condition after controlling for the included covariates. CUs in the POC sample were associated with 31 percent more entries than those in the RP sample. As expected, CUs with more than one member, higher income levels and higher education levels were associated with more entries. Of interest, if a CU provided all of their entries via total recall diaries, it was associated with a 73 percent reduction in the number of overall entries reported. If a FR risked having a CU not complete their diary (e.g., due to there being no entries when it was picked up), they would need to resort to collecting at least some of the CU's expenditures via recall and receipts in order to have a complete case for that diary week. Of note, these CUs made up a small proportion of the overall totals (2 and 9 percent in POC and RP samples, respectively). Also of interest, CUs who were eligible and chose to be placed with an online diary (for all participating CU members) were associated with a 48 percent reduction in the number of entries reported, controlling for other factors. This lower rate may be attributable to the login issues experienced by online diarists, though it could also be affected by other characteristics of these CUs (e.g. a younger CU composition). Model 2 shows that there were significant interactions between the sample condition and multi-member CUs and total recall diary CUs. Model 2 suggests that the higher entries for POC CUs could be attributable to the significantly larger number of entries among multi-member CUs (being in the POC condition controlling for this multi-member characteristic was no longer significant in this model). This is not surprising in light of the $20 incentive given for each diarist participating.

[25] See the Appendix for variable frequencies.

Although total recall diary CUs were associated with fewer entries in Model 1, these CUs did have a positive association with entries when CUs were participating in the POC compared to when they were in the RP sample providing total recall diaries.

c. Within-Household Reporting Characteristics

This subsection describes further research conducted into the patterns of diary reporting within households, mainly for POC CUs; research made possible by the provision of personal diaries in the POC test. The following aspects of diary reporting are covered:

- the extent of participation by various CU sizes,
- reporting levels and participation by demographic characteristics,
- participation based on member presence at diary placement, and
- reporting of 'sensitive' expenditures by test group and diary mode.

Participation by CU size

Of interest was how widespread completion of diaries was within POC sample CUs. This is shown below in Table 26, which indicates the extent of completion by the number of eligible CU members.

Table 26. Extent of CU completion by number of eligible diarists (N=325)[26]

CU completion	4-eligible CU (N=25)	3-eligible CU (N=62)	2-eligible CU (N=238)
Full	72%	63%	83%
Partial (3 members)	12%	--	--
Partial (2 members)	4%	29%	--
Partial (1 member)	12%	8%	17%
Total	100%	100%	100%

In CUs with 2 eligible members, 83 percent had full completion (i.e., both members either providing at least one entry or indicating they had no expenditures to report), and only 17 percent had partial, 1-member completion. It was seen that, across CUs of differing sizes, most CUs had all members who were eligible for the diary completing them. To the extent there was only partial completion, this was most prevalent in 3-eligible person CUs (37 percent of which had only 1 or 2 members completing their diaries).

[26] Of CUs with multiple eligible members and at least one diary completed; Includes total recall and legitimate blank diaries as completing.

We were further interested in whether certain respondents were responsible for providing the brunt of entries for their CU. To examine this, we ranked CUs members by the extent of diary entries they reported. The percent of overall CU entries reported by each member in a CU was calculated. The members were then ranked from highest to least according to the percent of entries that they supplied. Table 27 shows the average entries reported by ranked member.

Table 27. Ranking of member entries by number of eligible diarists (N=325)[19]

	4-eligible CU (N=25)	3-eligible CU (N=62)	2-eligible CU (N=238)
Most entries	31.3	41.0	31.8
Intermediate entries	12.4	--	--
Intermediate entries	5.4	12.5	--
Fewest entries	1.9	5.5	9.0
Average total	51.0	59.0	40.8

Using two-eligible-member CUs as an example, if the member providing more entries reported having 8 expenditures and the member providing fewer entries reported 2, and this was the same across all CUs, we would have 8 in the top right cell of Table 27 and 2 in the bottom right cell. In contrast, if both members reported equal numbers of expenditures (the case for only 5 percent of the 2-member CUs), say 5 each, both cells in that column would have the value of 5. Examining the actual data, members supplying the most entries provided on average 32 entries to the average of 9 entries reported by CU members with the fewest entries. Without access to the true average number of expenditures each member incurred, it is not possible to say whether these differences suggest underreporting among members supplying the fewest entries. However, these averages suggest a fairly uneven distribution of entries across members. Those providing the most entries in a CUs provided over 30 (or 64 to 78 percent of the total for those CUs). Conversely, the members supplying the fewest entries tended to supply 10 or fewer, varying by CU size. We examined these ‘low’ and ‘high’ reporters further in the next subsection.

Reporting and participation by demographic characteristics

We examined whether those who were eligible to provide diary entries varied in their participation by age (Table 28). Diarists providing no entries in their diary were distinguished from those with no expenditures to report (those with ‘legitimate blank diaries’).

[19] Of CUs with multiple eligible members and at least one diary completed; Includes total recall and legitimate blank diaries as completing.

Table 28. Distribution of diarist ages by extent of diary entries (among eligible to receive diary) (N=1,008)

Age	Overall (N=1,008)	no entries (N=105)	no expnses. (N=18)	1-3 entries (N=86)	4-20 entries (N=463)	21-50 entries (N=245)	51+ entries (N=91)	Mean entries
15-20 (N=110)	10.9%	19.1%	33.3%	29.1%	11.9%	1.6%	0.0%	5.5
21-34 (N=234)	23.2%	21.0%	22.2%	32.6%	24.4%	20.8%	17.6%	17.4
35-64 (N=495)	49.1%	41.9%	33.3%	32.6%	46.7%	56.7%	68.1%	23.6
65+ (N=169)	16.8%	18.1%	11.1%	5.8%	17.1%	20.8%	14.3%	20.6
Total	100%	100%	100%	100%	100%	100%	100%	

Table 28 shows that 49 percent of diarists were aged 35 to 64. These were also the 'high' reporters, making up 68 percent of those providing 51 or more entries – 24 on average. In contrast, the youngest eligible members were 'low' reporters. Making up 11 percent of the sample, a third of them reported that they had no expenses during the diary week, and a similar percent only supplied 1, 2, or 3 diary entries. Overall, this age group was disproportionately likely not to complete a diary (19 percent having no entries). These younger members were cited in past FR debriefings as likely to be the hardest to reach respondents, suggesting the need to motivate their participation. This notwithstanding, from a scan of expenditures among members with only 1 to 3 entries, it appears these younger diarists reported expenditures that likely would not have been captured had there been a design with only one household diary (e.g., video games, snacks, gasoline, and books).

Member presence at diary placement

Aside from age, another factor in the number of diary entries recorded was whether a CU member was present when diaries were placed (Table 29).

Table 29. Number of entries by respondent's presence at diary placement (among eligible diarists)[27] (N=1,007)

	N (members)	Mean	Median
Present	788	22.1	15
Absent	219	11.2	5

This analysis found that, when present at the FR's first visit, members reported more average entries in their diaries – 22 – than when they were not present – 11. As noted in the prior report, FRs were able to get a large proportion of members to attend diary placement, the case for 78 percent of eligible diarists[28].

[27] One CU was missing a value for presence/absence at placement. Note: may be attenuated by age as those present averaged 47 years old, those absent were on average 37 years old.
[28] FRs were trained in the importance of having members be present during placement, and the instrument for visit 1 included a prompt to gather other CU members to participate in the diary training.

Based on this finding it would be wise to reference the benefits of widespread within-household participation in upcoming diary placement protocols.

Reporting of sensitive expenditures

Additional analysis investigated whether the use of personal diaries in POC may lead to improved reporting of sensitive expenditures. The CE diary collects reports of whether respondents purchased alcohol with meals, and this was examined for POC and RP CUs (Table 30).

Table 30. Percent of CUs reporting alcohol expense with meals (and range of alcohol entries reported) by mode[29]

	N (CUs)	% w/alcohol
POC online (or mix)	190	27%
POC paper only	318	19%
POC online or paper	508	22%
RP	265	22%

There was no indication that POC sample CUs had more reports of alcohol than RP CUs – 22 percent of CUs in both samples reported some alcohol expenditure with meal purchases. However, this may be due to the large presence of paper diaries. POC diarists may have felt more comfortable reporting sensitive expenditures in a mode that could not be accessed by other CU members. When POC diarists used online diaries to report expenditures, 27 percent of CUs reported some alcohol expenditure during the one-week reporting period.

Diarist demographic characteristics by mode

Diarists had an option to choose between online or paper diaries (to the extent they were eligible to complete an online diary). We examined the demographic characteristics of individuals who chose between these modes. Table 31 shows the types of diary by respondent education level.

Table 31. POC diary type completed by CU member's education level (of those reporting education)

	Up to HS graduate	Some college and Associate	Bachelors	Masters and beyond	Total
Online	23.8%	28.1%	27.5%	20.6%	100%
Paper	42.7%	29.3%	20.3%	7.6%	100%

[29] Excludes 12 total recall CUs

There was a higher percent of CUs in which the member with the highest education level had a high school diploma or less, choosing a paper diary – 43 percent. CUs with members having higher education levels tended to opt for an online diary over a paper diary – 21 percent versus 8 percent. We also examined differences by age (Table 32).

Table 32. Diary type by age group

	15-19	20-39	40-64	65 and older	Total
Online	9.9%	43.5%	40.9%	5.8%	100%
Paper	8.2%	28.6%	40.4%	22.8%	100%

Among eligible diarists, those aged 20 to 39 were more likely to complete an online diary than a paper diary. In contrast, diarists age 65 and older were almost four times more likely to complete diaries using the paper mode. These findings support continuing the option of an online diary, as its preference will likely increase over time as the younger generation becomes a larger share of the total population.

VI. Conclusion

Interview expenditure reports were higher for the POC sample than for comparable production CUs, with income levels being similar across the two groups. This finding of significantly higher expenditure reports was not significant when controlling for CU characteristics and record use. The POC sample had significantly higher expenditures reported in the records interview for only the Rented Housing section, but had higher totals for two of the five sections in the recall interview. Higher record section reporting in the production sample may have been tied to the consistently higher expenditure amounts reported among these CUs when they used records compared to when they did not. Although POC CUs used records at high rates, record use by POC test condition did not reach significance in its association with higher expenditure amounts in the regression model. Record use may have been associated with other measures of data quality, however. POC CUs, on average, had fewer expenditure amounts that were rounded in the records interview sections. Furthermore, the POC sample had fewer expenditures requiring editing and lower missing data rates (for both expenditure and income reports).

Examining diary data, expenditure reports in the POC test were not significantly different than equivalent production amounts. Although POC CUs reported (non-significantly) lower median expenditure totals in the diary, this may have been due to lower income levels. POC CUs may have reported more small-value expenditures than RP CUs. For example, in the "All Other Products Services and Expenses" section, median POC amounts were lower than RP amounts, but the number of entries was higher. After controlling for other characteristics (e.g., income, education, household size), POC CU diarists were associated with a 31 percent greater number of expenditure reports than their RP counterparts. This was driven largely by more entries in multi-member CUs. Within-household analysis of the data indicate efforts to encourage presence at the diary placement and to obtain reports among younger diarists may help capture a greater extent of a CU's expenditures.

Overall, the analyses described in this report provide positive signs for the feasibility of the proposed survey redesign. Differential reporting by interview sections, and the associations between record use, expenditure totals, and the precision of expenditure amounts merit further examination. Efforts to encourage diary reporting among younger respondents and streamline the experience for those using online diaries may further improve upon the high quality of expenditure data observed in this POC test.

REFERENCES

Bagley, M., J. Edgar, L. Erhard, J. Gloster, T. Keller, P. Sarte, N. To, and C. Walker. Proof of Concept Test Final Report – Field Results. Bureau of Labor Statistics internal report. U.S. Bureau of Labor Statistics, Washington, DC, 2016.

Erhard, L., B. McBride, S. Park, and L. Tan. Proof of Concept Test for the Consumer Expenditure Survey: Results on Respondent Cooperation, Select Interview and Diary Characteristics, and Respondent Experience. Bureau of Labor Statistics internal report. U.S. Bureau of Labor Statistics, Washington, DC, 2016.

APPENDIX

METHODS

Method for deriving section-level totals

Using MTAB file, we subset to the EXPNAME in the record section rectypes. This was then limited to:

1) EXPNAME with positive mean totals, unless a negative mean represented a reimbursement (e.g., medical reimbursements),
2) expenditures with mean values in the range of $0 - $99,999 (i.e., to exclude the sale/purchase of property),
3) non-duplicates (e.g., vehicle payments being mapped to multiple UCCs, rent equivalence for homeowners, estimated interest amounts)

…resulting in the variables used for analysis*

NOTE: the totals represented 3-months' worth of expenditures for each section

We then summed the values to the section level, and accounted for CUs with no expenditures in a specific section by creating 0-values. 3 RP CUs had no expenditures reported in any of these categories (e.g., had outcome dispositions '203'/transmit with no more follow-up possible).

*EXPNAMEs in POC record sections (reported by either RP or POC CUs):

2) JRTPAYQV QRT3MCMX
3) QADPTAX QADINT1X QADINT2X QADINT3X PDAMTX1 PDAMTX2 PDAMTX3 JFEETOTX QLR3MCMX QSPASSX QSPCLX
4) QADAPG1X QADAPG2X QADAPG3X QADCAB1X QADCAB2X QADCAB3X QADINE1X QADINE2X QADINE3X TELCEL1X TELCEL2X TELCEL3X TELRES1X TELRES2X TELRES3X QFCD3MCX INTCHGX QADFUL1X QADFUL2X QADFUL3X
10) QADRENTX QADDOWNX QADFEEX QADPMT1X QADPMT2X QADPMT3X QEXTRA1X QEXTRA2X QEXTRA3X TRADEEXP
11) EXREIMBX QADITR1X QADITR2X QADITR3X QDNPYMTX QTRADEX
13) QLIFCMX QLNGTCMX QOTHCMX QPR3MCMX QTN3MCMX QVH3MCMX
14) QHI3MCX QCUMED1X QCUMED2X QCUMED3X QDRGPM1X QDRGPM2X QDRGPM3X
15) MEDPMTX MEDRMBX

We repeated the process above for EXPNAME in recall section rectypes. Unlike for the record sections, no variables were excluded* when we applied the below criteria:

1) variables with positive mean totals (all were positive)
2) expenditures with mean values in range of $0 - $99,999 (nothing excluded)
3) exclude duplicates (no duplicates noted)

We again summed values at the section level, and accounted for CUs with no expenditures in a specific section. 41 POC CUs and 167 RP CUs did not report expenditures for any of these variables.

*EXPNAMEs in POC recall sections (reported by either RP or POC CUs):

6) INSTLLEX MAJINSTX MAJPURX MAJRENTX MINPURX MINRENTX
8) FURNPURX QFRT3MCX QREP3MCX
17) SUBEXPX
18) TOTYUPDX CMBUSX CMLOCALX CMPLANEX CMSHIPX CMTRAINX GASOILX LDGCOSTX PARKINGX QTRFLAX QTRGLAX RTBOATX RTCARX RTCAMPX RTOTHERX TRMISCX TRPALCGX TRPALCHX TRPETRTX TRPGFTCX TRPSPRTX TRPTOLLX TRNONCUX ALCGROCX ALCMEALX LOCADMSX LOCLODGX QLCGLAX QLCMLAX
19) MISCEXPX CONTEXPX

Method for deriving count of interview expenditure entries

Subset expenditure records (in post-phase 3 MTAB file) to those without allocation or only manual updates without adjustments (e.g., ALCNO=0, 1, 100)

Identify unique expenditure counts by removing duplicate records (e.g., in the case of months; sub-setting to a unique SEQNO). This also avoids duplication across separate counts for principal and interest payments, or property values and rental equivalences.

- For certain CE interview sections, count each record associated with a SEQNO (e.g., in the case of rectypes XPA & XPB/Section 20, take each unique expenditure (EXPNAME) within a SEQNO). Same for Sect 18 RTYPEs
- For rectype XPA/Section 20A, exclude reports that are used to subtract from other reports to obtain an amount (e.g., grocery purchases minus those not for food and non-alcoholic beverages)

Add up total counts at section level after duplication. Assign zero values for sections without any expenditure counts.

Description of Branch of Production and Control (P&C) Data Processing Procedures

For the POC interviews and diary, P&C followed a number of procedures to ensure the quality of the collected responses:

- Imputation was used to adjust for missing values, although income imputation was not carried out, unlike when processing production data.
- Allocation involved costs for combined expenditures being assigned to various component items.
- Summary variables were created, allowing for higher-level expenditure analysis.
- Outlier detection and checks for item misclassification was carried out. Outlier detection was done at the UCC level, leading to some adjustments (e.g., flagging high amounts for leased vehicles attributable to a 1-time lease payment). There was a limit on the outlier detection that was possible for the interview data given the absence of prior-interview values for comparison. Additionally, the absence of a paper diary form for all diary cases hampered the ability to detect unexpected values, as is possible with production diary review.
- Mapping to UCCs occurred, but for the interview was limited by the lack of UCF data; this limitation required statistical methods staff to carry out manual weight adjustments.
- For the POC test diary, production processes resulted in an (empty) second week of data being created, as is done with production data structures. This additional week was ultimately excluded when analyzing the test data.

Description of Statistical Tests

To determine whether or not differences in the report were statistically significant, we used Wilcoxon-Mann-Whitney (non-parametric) tests. These were used as the data were not normally distributed, an assumption of the independent samples t-test. The Wilcoxon-Mann-Whitney test measures the probability that a randomly selected expenditure from one sample is greater than or less than a randomly selected expenditure from another sample. The probability "p" of a randomly selected expenditure from one sample being greater than a randomly selected expenditure from another sample can be estimated from the data. Under the null hypothesis that the expenditures have the same distribution, the probability "p" is a random variable with an expected value of one-half and a variance of $\sigma^2 = (n_{s1}+n_{s2}+1) / (12n_{s1}n_{s2})$, where n_{s1} and n_{s2} are the number of reported records from the compared samples. This leads to the Wilcoxon-Mann-Whitney test statistic being defined as a z-score:

$$z = (p - ½) / \sigma.$$

When the sample sizes are large enough this z-score has a standard normal distribution. The computer program procedure used for the test (SAS' NPAR1WAY) incorporates a continuity correction when computing the standardized test statistic by subtracting 0.5 from the numerator if it is greater than zero. If the numerator is less than zero, PROC NPAR1WAY adds 0.5. Some sources recommend a continuity correction for nonparametric tests that use a continuous distribution to approximate a discrete distribution. For significance findings, we reported two-sized test probabilities using a continuity correction to be more conservative in our findings. The first diary regression model and first income regression model used a stepwise selection method to identify significant variables at a p=0.05 probability level. Some diagnostic testing was run on the first model to detect if certain CUs had high leverage over the model findings and whether model assumptions were met. Additional diagnostic tests were run for the income regressions.

DATA TABLES

Diary Data Quality

Total amount of expenditures reported per CU, overall with total recall included

	N	Mean	Median
RP	283	$974.73	$510.64
POC	520	$1,085.97	$524.40

Total amount of expenditures reported per CU, by section with total recall included

	POC (N=520)		RP (N=283)		
	Mean	Median	Mean	Median	Wilcoxon Z
Z_FDB	$109.52	$75.81	$110.71	$79.67	0.1290
Z_MLS	$80.47	$49.00	$70.82	$40.33	-2.6647*
Z_CLO	$42.87	$0	$47.59	$0	-2.4639*
Z_OTH	$853.10	$272.28	$745.61	$339.61	0.3073

*p<.05; **p<.01; ***p<0.001

Total amount of expenditures reported per CU, overall and by section excluding extreme values

	POC (N=506)		RP(N=258)		
Variable	Mean	Median	Mean	Median	Wilcoxon Z
ZTOTAL	$917.65	$524.40	$979.80	$578.34	1.107
Z_FDB	$110.05	$76.22	$111.94	$82.46	0.3363
Z_MLS	$80.78	$49.09	$75.31	$45.59	-1.7553
Z_CLO	$43.40	$0	$51.52	$0	-1.6516
Z_OTH	$683.42	$272.28	$741.03	$381.16	1.3522

Interview income distributions

Income (FINCBTAX) quintile

	20th percentile	40th percentile	60th percentile	80th percentile
RP	$12,000	$32,000	$58,600	$106,000
POC	$14,553	$32,584	$65,000	$113,800

Interview regression variable values

Variable	Value	Label	Frequency	%
Test condition	0	RP sample	1,483	74%
	1	POC sample	520	26%
			2,003	100%
Number of members (CUSIZE)	0	Single member	615	31%
	1	Multiple members	1,388	69%
			2,003	100%
At Least Some Record Use (RECORDS for RP, S2RECRD-S22RECRD for POC*)	.	Missing	6	0%
	0	Less than 10%	893	45%
	1	More than 10%	1,104	55%
			2003	100%
Imputed post-tax family income (FINCBTAX)	0	=<$0	162	8%
	1	$1-$25,000	533	27%
	2	$25,001-$58,000	488	24%
	3	$58,001-$110,000	432	22%
	4	$110,001-above	388	19%
			803	100%

Highest education among members (EDUCA)	1	HS grad or less	587	29%
	2	Some college	423	21%
	3	Assoc./bachelor's degree	691	35%
	4	Advanced degree	302	15%
			2,003	100%
Homeowner (CUTENURE)	0	Owned	763	38%
	1	Rented/student housing	1240	62%
			2,003	100%
Race of respondent (MEMBRACE)	0	White	1,604	80%
	1	All other race	399	20%
			2,003	100%
All Valid Blanks (FINCBTAX)	0	At least one income source is missing	141	7%
	1	All valid entry	1,862	93%
			2,003	100%

*In POC, record use questions were asked at the end of each selected-section (S2RECRD S3RECRD S4RECRD S10RECRD S11RECRD S13RECRD S14RECRD S15RECRD S21RECRD and S22RECRD). In order to determine the frequency of records use in POC, we used the proportion of sections having record use reported (the number of reported sections with records used/total number of sections where record use was asked about) as the ratio to compare with the RP record frequency (RECORDS) question. Essentially, a value ‘less than 10%’ in POC indicated no record use, for which the closest comparable category in RP was record use ‘Never or almost never (less than 10% of the time).’

Diary regression variable values

Variable	Value	Label	Frequency	%
Test condition	0	RP sample	283	35%
	1	POC sample	520	65%
			803	100%
Number of members (CUSIZE)	0	Single member	250	31%
	1	Multiple members	553	69%
			803	100%

Whether HH only provided entries via total recall	0	HH with some direct entries	767	96%
	1	HH with total recall diaries	36	4%
			803	100%
Imputed post-tax family income (FINCBEFX)	0	$0	84	10%
	1	$1-$25,000	213	27%
	2	$25,001-$58,000	173	22%
	3	$58,001-$110,000	162	20%
	4	$110,001-above	171	21%
			803	100%
HH with all entries coming via the web (POC only)	0	Some paper	655	82%
	1	All web	148	18%
			803	100%
Highest education among members (HIGH_EDU)	1	HS grad or less	171	21%
	2	Some college	164	20%
	3	Assoc./bachelor's degree	302	38%
	4	Advanced degree	166	21%
			803	100%