

## ARTICLES

# Availability of Respondent Provided Financial Records to Supplement Self-Reports of Household Expenditures

Emily Geisen<sup>\*</sup>, Charles Lau<sup>†</sup>, M. Christopher Stringer<sup>‡</sup>, Jennifer Edgar<sup>\*\*</sup>, Brandon Kopp<sup>††</sup>, Ashley Richards<sup>‡‡</sup>

Keywords: survey practice

<https://doi.org/10.29115/SP-2016-0031>

---

## Survey Practice

Vol. 9, Issue 5, 2016

---

Use of respondent records is a method for supplementing survey data. Records can improve data quality when data collected by self-report are subject to recall or other measurement errors. One limitation of using respondent records is that respondents often cannot provide records for all survey items of interest. We explored the availability of respondent records using data from the Consumer Expenditure Records Study (CE Records Study). The CE Records Study is a non-probability feasibility study that examined the accuracy of self-reported data for various types of household expenditures. In the first interview, participants provided self-reports about the cost of household expenditures from the previous three months. In a follow-up interview, participants provided records (e.g., receipts, bank statements, bills) for all expenditures asked about. By comparing self-reports and records, we were able to evaluate the accuracy of self-reports. Records were available for 36% of the 3,039 expenditures reported. There were several expenditure and respondent characteristics that were associated with the availability of records suggesting that use of retrospective record collection may be more successful for supplementing self-reported data for certain types of consumer surveys compared to others.

---

## INTRODUCTION

Retrospective self-reports are the main method of collecting data about consumer expenditures. However, self-reported data are often subject to recall error, which can impact the quality of survey data (Groves 1989). One way to reduce recall error in surveys of consumer expenditures is to supplement self-reported data through the use of financial records provided by respondents, such as store receipts, utility bills, or bank or credit card statements (Edgar and Gonzalez 2009; Laurie and Moon 2010; Kashiara and Wobus 2006; Safir and Goldenberg 2008).

---

\* RTI International

† RTI International

‡ RTI International

\*\* U.S. Bureau of Labor Statistics

†† U.S. Bureau of Labor Statistics

‡‡ RTI International

Respondents' financial records can be collected in one of two ways: prospectively or retrospectively. If collected prospectively, researchers ask respondents in advance to save records for expenditures they make during the survey reference period. If collected retrospectively, respondents are asked at the time of the survey to retrieve any available records for expenditures they made during the survey reference period. Although prospective record collection may yield more records, asking respondents to collect records may reduce response rates (Couper, Ofstedal, and Sunghee 2013). It is also possible that asking for records may alter respondents' purchasing behavior, possibly introducing bias. Retrospective record collection avoids these problems by having respondents gather records for purchases already made. However, one challenge is that respondents might not have records for all past expenditures. In this paper, we evaluate whether retrospectively collected financial records can be used to supplement self-reports by assessing the extent to which respondents were able to provide retrospective records for the expenditures asked about in a survey of consumer expenditures. We also investigate how the availability of records was associated with the type of expenditure (e.g. clothing, furniture, telephone bills) and respondent characteristics (e.g. age, income, sex).

## **DATA AND METHODS**

We analyzed data from the Consumer Expenditure (CE) Records Study, which was designed as a feasibility study for measuring accuracy of self-reports in the Consumer Expenditure Quarterly Interview Survey (CEQ). The CEQ, combined with a separate diary survey, is the primary source of information about personal expenditures in the United States. The U.S. Census Bureau contracted with RTI International (RTI) to conduct the CE Records Study. RTI collected data on 3,039 expenditures reported by 115 respondents. Respondents were recruited through convenience sampling methods in Raleigh-Durham, NC and the Washington, D.C. metropolitan area. Respondents completed two interviews, spaced 4–7 days apart, and were provided a \$40 cash incentive to complete Interview 1 and a \$60 cash incentive to complete Interview 2. Data were collected via computer-assisted personal interviewing (CAPI) in respondents' homes between February and May 2011.

In Interview 1, respondents completed an abbreviated version of the CEQ instrument; respondents provided self-reports about seven types of expenditures (housing, utilities, appliances, furniture, clothing, health insurance, and miscellaneous). For each expenditure type, respondents reported whether their household had made an expenditure in the past 3 months and if so, the amount of the purchase. The question wording, interviewer instructions, and other data collection procedures in the CE Records Study were nearly identical to the procedures used by the U.S. Census Bureau during standard CEQ interviews<sup>1</sup>.

At the end of Interview 1, respondents were asked to collect financial records for the expenditures they reported in the interview. We defined “financial records” broadly to include receipts, bills, bank and credit card statements, and respondents’ notes or budgets in either electronic or hard copy form. Four to seven days after Interview 1, the second interview occurred, during which the respondent provided the records to the interviewer. When a record was not available for an expenditure reported in Interview 1, respondents were asked why. The amount on the record was compared to the self-report from Interview 1. If there was a difference, interviewers asked respondents questions to identify possible reasons for the discrepancy. This qualitative information was used to better understand the reasons why records were unavailable or why the self-reports were inaccurate.

## RESULTS

Out of the 115 respondents, 106 provided records for at least one expenditure reported in Interview 1. Out of 3,039 expenditures reported in Interview 1, respondents provided records for 1,082, which is 36 percent of reported expenditures. Table 1 shows the percent of expenditures with a record for each respondent and expenditure category.

The results in Table 1 suggest that there is substantial variation in the availability of financial records, particularly by respondent characteristics (race, housing status, income) and the type of expenditure. To control for confounding variables, we estimated a logistic regression that predicted record availability based on the characteristics from Table 1. Because there is likely a correlation between expenditures reported by the same respondent, we adjusted the model for the clustering of expenditures within respondents. Table 2 shows the parameters from the logistic regression. Reference categories are listed in parentheses.

The results in Table 2 show that records were more likely to be available from respondents who were non-Hispanic White, homeowners, highly educated, from smaller households, or had higher income. Expenditures that were purchased more recently – within 1 month prior to the interview compared to 2 or 3 months prior to the interview – were more likely to have a record. Expenditures that were more expensive were more likely to have a record. In addition, expenditures that tended to be recurring such as expenditures for housing (e.g. mortgage and rent); phone/Internet (e.g. monthly phone, Internet, and cable); or utilities (e.g. water, gas, electric) were more likely to have a record compared to purchases of appliances (e.g. toasters, coffee machines, dishwashers) or furniture (e.g. lamps, curtains, coffee tables).

---

1 The Consumer Expenditure Surveys Quarterly Interview CAPI Survey (2011-12) is available at: <http://www.bls.gov/cex/capi/2011/cecapihome.htm>.

**Table 1** Record availability by respondent and expenditure characteristics (n=3,039).

Respondent characteristics	Items reported	% with record	Expenditure characteristics	Items reported	% with record
Gender			Amount		
Woman	1,963	38%	\$1–\$29	814	30%
Man	1,076	32%	\$30–\$74	878	37%
Age			\$75–\$149	640	42%
18–34	977	33%	\$150 or higher	610	40%
35–59	1,287	37%	Type		
60 and over	775	37%	Housing	260	39%
Race-ethnicity			Phone/Internet	572	38%
Hispanic or non-White	935	20%	Utilities	613	37%
Non-Hispanic White	2,104	43%	Appliances	300	24%
Educational attainment			Furniture	225	25%
High school or less	578	28%	Clothing	590	38%
Some college	890	35%	Miscellaneous	479	38%
College or higher	1,571	39%	Time since expenditure		
Employment status			Current or last month	797	46%
Not employed	1,418	35%	2–3 months prior	1,361	38%
Employed	1,621	36%	Not applicable <sup>a</sup>	881	23%
Household income					
Less than \$30,000	888	26%			
\$30,000–\$60,000	1,005	38%			
More than \$60,000	1,146	41%			
Household size					
More than one person	2,127	37%			
One person	912	33%			
Location					
North Carolina	2,721	35%			
Washington, D.C. area	318	43%			
Housing status					
Owner	1,795	41%			
Renter	1,244	28%			

<sup>a</sup>Time since expenditure not available for certain types of expenditure categories.

There were two significant interactions: one between age and education and one between household size and income. For respondents with a low level of education, older respondents were more likely to have a record than young respondents. For respondents with a medium or high level of education, age did not have a statistically significant association. In addition, while household size was positively associated with record availability for low or medium income respondents, there was no association between household size and record availability for high-income households.

**Table 2** Logistic regression analysis predicting availability of records (n=3,039).

Predictor	$\beta$	SE	P
Intercept	0.74	0.70	0.29
Respondent characteristics			
Female (vs. male)	0.32	0.21	0.13
Age (in years)	-0.01	0.01	0.19
Non-Hispanic White (vs. other)	0.51	0.25	0.04
High school or less (vs. college)	-2.48	0.98	0.01
Some college (vs. college)	-1.11	0.68	0.10
Working (vs. not working)	0.09	0.21	0.67
D.C. (vs North Carolina)	-0.46	0.31	0.14
Household size	-0.32	0.11	<0.01
Low income (vs. high)	-0.95	0.55	0.08
Medium income (vs. high)	-1.30	0.53	0.01
Owners (vs. renters)	0.58	0.27	0.03
Expenditure characteristics			
Housing ( vs. appliances)	0.75	0.30	0.01
Phone/Internet (vs. appliances)	0.63	0.22	0.00
Utilities (vs. appliances)	0.55	0.24	0.02
Furniture (vs. appliances)	-0.15	0.24	0.52
Clothing (vs. appliances)	1.76	0.29	<0.01
Miscellaneous (vs. appliances)	1.67	0.29	<0.01
Housing ( vs. appliances)	0.75	0.30	0.01
Recent (vs. not recent)	0.33	0.10	<0.01
Interactions			
Age x high school or less	0.05	0.02	0.02
Age x some college	0.02	0.01	0.15
Household size x low income	0.30	0.17	0.07
Household size x med income	0.42	0.15	0.01
QIC Goodness of fit	3428.6		

QIC = Quasilikelihood under the Independence model Criterion.

## DISCUSSION

In Interview 2, respondents were only able to provide records for 36 percent of the expenditures they reported in Interview 1. Although record availability was limited overall, a recent National Academies of Science report on the CEQ concluded that “The use of records is extremely important to reporting expenditures and income accurately” (Dillman and House 2012, 75). Therefore, even a limited set of records may substantially improve data quality compared to self-reports alone. Furthermore, there were several expenditure and respondent characteristics that were associated with the availability of records suggesting that use of retrospective record collection may be more successful for supplementing self-reported data for certain types of consumer surveys compared to others.

In particular, expenditures that were more expensive, more recently purchased, or recurring were more likely to have records. Qualitative data collected during the study suggested that respondents did not keep records for inexpensive items or items they were not planning to return. This suggests that retrospective records may work better for capturing data about significant expenditures such as those over \$200. However, other methods for boosting data quality, such as prospectively asking for records, may be needed to capture information about smaller, everyday purchases.

Respondents also indicated that they did not tend to keep certain records very long. The CEQ uses a 3-month reference period, which partially explains why so few records were available for expenditures reported in the survey. Retrospective records may work better for surveys with shorter reference periods such as 1 or 2 weeks compared to 3 months.

Finally, respondents were more likely to have records for expenditures that were purchased or paid for on a recurring basis, such as rent/mortgage, phone, and utilities. For many of these types of recurring records, respondents were able to look up the information online or retrieve it from their email even if they did not actively save or keep the information. This suggests that retrospective records may work well for capturing data about recurring expenditures such as utilities.

Record availability was also related to race and several respondent characteristics associated with having a higher socioeconomic status: higher education, higher income, and home ownership. Across the expenditure categories, respondents who were non-Hispanic White and had a higher socioeconomic status were more likely to have records for their purchases.

While our results on the use of respondent records in expenditure studies are informative, our conclusions are limited by the fact that the CE Records Study is a small-scale study based on a convenience sample in two geographic areas. Future research based on larger probability samples is needed to provide stronger recommendations about the use of respondent records for supplementing self-reports. Additional studies should compare asking respondents to collect records retrospectively versus prospectively. While the use of prospective record keeping may yield more records, it is unclear if this would affect respondent behavior. For example, respondents may purchase fewer items or spend less money if they are tracking their expenditures. Respondents may purchase different types of items knowing that they would have to share more detailed information about their purchases with researchers.

Respondent records have the potential to play an important role in the collection of expenditure data. Although only 36 percent of expenditures reported had records, we found cause for significant concern in the accuracy of self-reported data (Geisen et al. 2012). Only 30 percent of respondents'

self-reports of the costs of the expenditure matched the records exactly. Respondents tended to underestimate and overestimate costs in equal proportions. Comparing the absolute difference in cost between respondents' self-reports and records, respondents misreported the amount of items they purchased by 30 percent. For example, the respondent reported the cost as \$100, but the record showed \$130.

Incorporating respondent records into a survey design has the potential to yield more accurate data, while reducing the reporting burden on the respondent. While record availability was found to vary by respondent and expenditure characteristics, additional research on how to improve the number of respondent records provided could prove to be invaluable for surveys collecting this type of detailed information.

These findings suggest that the use of retrospective record collection to supplement self-reported survey data would be most successful in surveys that use a short reference period, ask about significant or recurring expenditures, and in surveys that target populations with higher socioeconomic status.

## REFERENCES

- Couper, M., M.B. Ofstedal, and L. Sunghee. 2013. "Encouraging Record Use for Financial Asset Questions in a Web Survey." *Journal of Survey and Statistical Methodology* 1 (2): 171–82.
- Dillman, D.A., and C.C. House, eds. 2012. "Panel on Redesigning the BLS Consumer Expenditure Surveys." Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. [http://www.nap.edu/catalog.php?record\\_id=13520](http://www.nap.edu/catalog.php?record_id=13520).
- Edgar, J., and J.M. Gonzalez. 2009. "Correlates of Data Quality in the Consumer Expenditure Quarterly Interview Survey." In *Proceedings of the American Statistical Association A*, 1151–63.
- Geisen, E., C.Q. Strohm, B. Kopp, and A.K. Richards. 2012. "The Relationship between the Accuracy of Self-Reported Data and the Availability of Respondent Financial Records." Presented at Annual Conference of the American Association for Public Opinion Research, Orlando, FL.
- Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Kashihara, D., and D. Wobus. 2006. "Accuracy of Household-Reported Expenditure Data in the Medical Expenditure Panel Survey." In *Proceedings of the American Statistical Association*, 3193–3200.
- Laurie, H., and N. Moon. 2010. "Is It Worth the Effort?" Presented at the Annual Meeting of the American Association for Public Opinion Research, Chicago, IL.
- Safir, A., and K.L. Goldenberg. 2008. "Mode Effects in a Survey of Consumer Expenditures." In *Proceedings of the American Statistical Association, Alexandria, VA*, 4336–43.