

Using Contact Attempt History Data to Determine the Optimal Number of Contact Attempts ⁺

Adam Safir¹ and Lucilla Tan²

¹U.S. Bureau of Labor Statistics, Washington, DC, safir.adam@bls.gov

²U.S. Bureau of Labor Statistics, Washington, DC, tan.lucilla@bls.gov

Abstract

The Consumer Expenditure Quarterly Interview Survey (CEQ) expends a disproportionate amount of effort attempting to collect data from difficult-to-interview respondents. Maintaining acceptable response rates, improving sample representativeness, and reducing potential nonresponse bias are factors that drive multiple contact attempts. However, data from difficult-to-interview respondents come at a relatively high cost. In the face of budgetary constraints, a decision to lower the maximum number of contact attempts is one way to reduce overall data collection costs. In this study, we attempt to determine an optimal contact attempt threshold by evaluating the tradeoffs of four fundamental survey performance measures along the continuum of contact attempts: survey response rate, sample representativeness, reporting quality, and cost.

Keywords: Contact history, data quality, field operations, level of effort, and paradata.

1. Introduction

In a constrained budgetary environment, the challenge of collecting high quality data while still controlling costs takes on a heightened priority. Attempts to contact eligible respondents are especially costly for personal visit surveys, and in particular for surveys with eligible sample units that are dispersed across a large geographic area. A natural question for survey producers interested in better managing data collection costs is how to determine the “optimal” number of attempts to complete an interview.

In our attempt to answer this question, we explore an approach that gives consideration to four fundamental survey performance measures: survey response rate, sample representativeness, reporting quality (as a dimension of data quality), and cost. Changing the number of contact attempts will have differential impacts on these measures: for example, decreasing the number of contact attempts will reduce collection costs but it will also adversely affect response rates, while the effects on data quality and sample representativeness may not be immediately known.

A number of studies have investigated the manner in which high effort cases contribute to nonresponse bias in surveys. These studies are based on the “continuum of resistance” paradigm, in which sample units are ordered by the number of contact attempts required to complete the interview. The paradigm assumes that respondents requiring higher effort to contact are more similar to true nonrespondents (Groves 2006). However, Heerwegh et al. (2007), Srinath et al. (2003), Bates and Creighton (2000), Curtin et al. (2000), Keeter et al. (2000), and Dennis et al. (1999) found that in general, absolute differences in key survey estimates across level of effort groups were small. In addition, although

⁺ The opinions expressed in this paper are those of the authors and do not represent official policy of the Bureau of Labor Statistics.

demographic characteristics may differ between respondents requiring high and low contact attempt effort, the extent of these differences vary along the continuum of contact attempts. In other words, respondents requiring slightly different levels of effort were more different in the lower part of the continuum, but more similar in the upper part. This suggests that the strategy of increasing response rates at the higher end of the continuum of attempts may not be as successful as initially thought, in terms of drawing in respondents dramatically different from those already in sample, or more similar to true nonrespondents (Curtin et al. 2000).

Furthermore, Heerwegh et al. (2007) found the composition of nonresponse error to be relatively stable, with noncontact making the largest contribution to nonresponse error, even when refusal rates were higher than noncontact rates. This helps to explain why the effort to increase response rates through extended contact attempts may not always translate into a large reduction in nonresponse error, particularly if residual nonrespondents behave differently from completed respondents in ways that are germane to key survey estimates. Therefore, if the purpose of extended contact attempts is to increase response rates (at sometimes great cost) in order to minimize the potential for nonresponse bias, this brief survey of the literature indicates that higher response rates *alone* may not be the most efficient approach. Nonetheless, some studies advocate pursuing high effort cases because even if not completely reflective of nonrespondents, data from these cases may help to minimize non-sampling error and potential bias in survey estimates.

1.1 The Quarterly Consumer Expenditure Interview Survey (CEQ)

The Bureau of Labor Statistics sponsors the collection of data on spending by America's consumers through the Consumer Expenditure Survey (CE) program. The CE program consists of two separate and independent surveys: the Diary Survey and the Interview Survey. Data collection for both surveys is conducted by the U.S. Census Bureau. The Diary Survey, designed to collect small expenditures that would be difficult for respondents to recall during a long reference period, is a self-administered survey in which respondents record all household expenditures for two one-week periods. The Interview Survey (CEQ) is a panel survey conducted over five consecutive calendar quarters and is designed to collect larger purchases that a respondent could reasonably be expected to recall for a period of three months. Each of the five quarterly interviews is referred to as a "wave" of data collection. The CEQ is a computer assisted personal interview (CAPI) survey, although telephone interviews do occur. Results from the Diary and Interview surveys are integrated to create official expenditures estimates.

The CE program introduced a Contact History Instrument (CHI) to the CEQ in April 2005. The CHI module enables interviewers to maintain detailed information about each contact attempt for their assigned cases, such as day and time of the contact attempt, outcome of the attempt, strategies used to attempt contact, and perceived concerns that respondents have about participating in the survey.

1.2 Analysis Data

This study uses CEQ and CHI data from the period of April 2006 through March 2008. We base our recommendation for the optimal contact attempt threshold on an analysis of Wave 1 cases. We use Wave 1 data for the main analysis for several reasons. As the "initial" interview, these cases (1) generally require greater effort than subsequent panel interviews, (2) are "uncontaminated" by previously completed interviews, and (3) are primarily conducted by personal visit. The Wave 1 response rate for the study period was

76.0 percent out of 16,897 eligible cases. The response rate for the final Wave 1 sample used in the analysis was 70.6 percent, out of 13,792 eligible cases.¹ We subsequently used Wave 2 completed interviews ($n=13,742$) to evaluate the impact of our recommended optimal number of contact attempts on key aggregate survey estimates.

About 85 percent of the eligible sample addresses have been resolved by the seventh contact attempt. Among completed interviews, 85 percent are resolved by the sixth contact attempt (Figure 1); after the sixth attempt, the marginal, or incremental, gains fall under five percent.

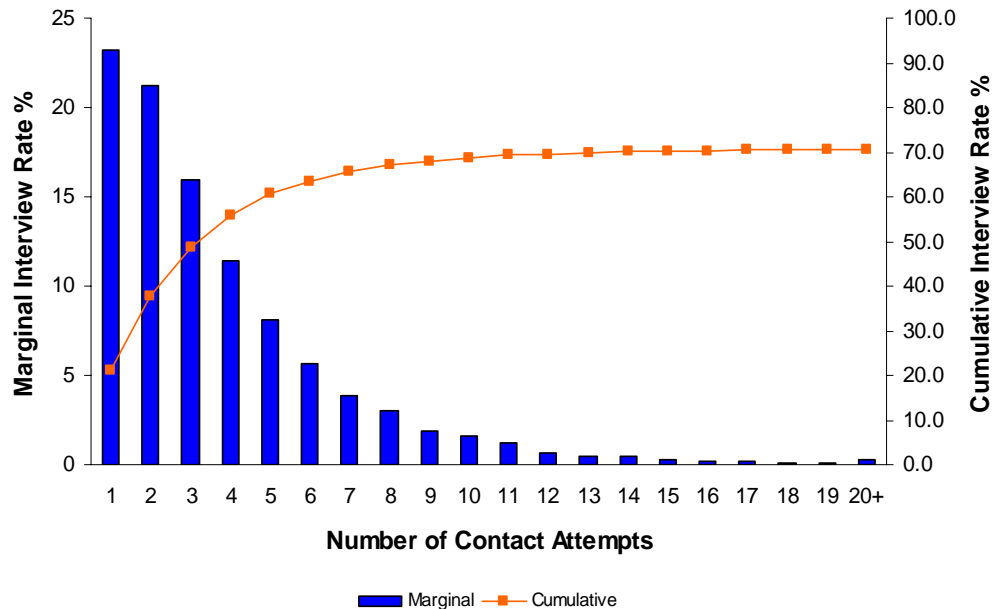


Figure 1: Interview Rate by Contact Attempt

2. Methods

2.1 Formation of Comparison Groups

The number of contact attempts for completed interviews range from 1 to more than 40; therefore we used cluster analysis to form a smaller set of aggregated comparison groups. Based on the results of scatter plot and regression analyses showing reporting quality to be associated with the number of contact attempts and level of expenditures (see Table 1), we included the following inputs in a cluster analysis to form comparison groups:

Completeness and Reporting Accuracy

- Percent of contact attempts by personal visit
- Percent of expenditure questions answered with “don’t know/refused”
- Length of interview (minutes)

¹ The final Wave 1 sample used in the analysis was the result of dropping of eligible cases that did not meet logical consistency checks among interviewer perceived concerns about the respondent, a key input variable in our analysis. However, we also reanalyzed the Wave 1 data, including all Wave 1 cases in the study period, and found no changes to the findings presented in this paper.

Completeness and Reporting Accuracy, contd.

- Use of recall aids such as the information booklet or records/receipts

Respondent Cooperation

- Number of soft refusals in the respondent’s contact attempt history
- Perceived Concerns Index (PCI)

The PCI is a summary index we constructed from the interviewer’s “doorstep” perception of respondent concerns regarding survey participation recorded in the CHI. Recent research on these doorstep concerns has found these measures to be good predictors of survey nonresponse (Bates et al. 2008), and also to be associated with reporting quality (Tan and Tsai 2008). The PCI was constructed using principal component analysis on the CHI concern options that the interviewer can mark off at each contact attempt. The three retained principal components used in the construction of the PCI accounted for 68 percent of the total variance of the perceived concerns, and relate to perceived respondent concerns about privacy, survey content, time, and respondent hostility. The PCI is an ordinal index, in which a higher PCI value indicates the respondent is perceived to have more of these concerns than another respondent with a lower PCI value.

Table 1: Examining Associations of Reporting Quality Indicators by Number of Contact Attempts and Reported Expenditures⁺

Covariate	Dependent Variable: Unedited Expenditures			Dependent Variable: Number of Contact Attempts		
	Parameter Estimate	SE	P-value	Parameter Estimate	SE	P-value
Intercept	889.2	736.4	0.2273	4.74	0.12	<.0001
Perceived Concerns Index	-151.9	50.6	0.0027	0.12	0.01	<.0001
Use Information Book	166.3	343.3	0.6282	-0.20	0.06	0.0002
Use Records	226.3	357.2	0.5265	-0.21	0.06	0.0002
Percent by Personal Visit	-1,665.3	696.0	0.0167	-2.62	0.11	<.0001
Number of Soft Refusals	-61.6	706.9	0.9305	1.07	0.11	<.0001
Percent of Exp. Questions with “Don’t Know/Ref”	-1.9	15.7	0.9055	0.00	0.00	0.5202
Length of Interview (min)	64.9	6.1	<.0001	0.02	0.00	<.0001
Size of CU	762.9	112.4	<.0001	-0.05	0.02	0.0056
<i>N</i>	9,681			9,700		
<i>R</i> ²	0.023			0.16		

⁺ Wave 1 data

In the cluster analysis, we used hierarchical clustering with Ward's minimum variance approach as the measure of dissimilarity to form the analysis groups (Khattree and Naik 2000). The clustering history is shown in Figure 2. Since the clusters are linearly related to the number of contact attempts, we can form meaningful groups based on the number of contact attempts. Based on sample size considerations and our judgment on the practicality and feasibility of the number of groups, we decided on three comparison groups: one to four attempts, five to seven attempts, and eight or more attempts.

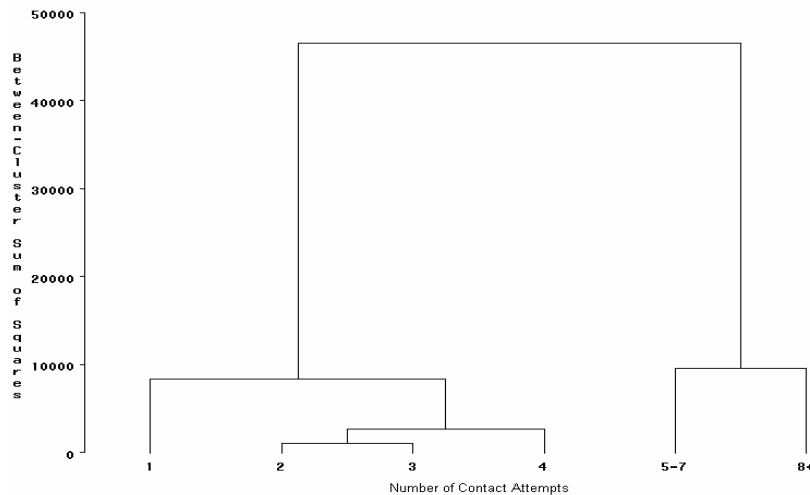


Figure 2: Clustering History and Formation of Analysis Groups

2.2 Reporting Quality Index (RQI)

To facilitate the comparison of the reporting quality characteristics by contact attempt group, we created a summary index of reporting quality from all reporting quality indicators used in the cluster analysis. These variables were a mix of binary and continuous variables, so we first used principal components analysis to construct an Interim Reporting Quality Index (IRQI) based on the following variables: percent of contact attempts by personal visit, percent of don't know/refused responses, length of interview, use of recall aids, and number of soft refusals in the respondent contact attempt history. This set of variables excludes the interviewer's perception of respondent concerns since those measures were previously used in the construction of the PCI.

We coded all IRQI variables so that higher values indicate better reporting behavior. The two retained principal components for the IRQI accounted for 48 percent of the total variance of the five reporting quality variables, and were related to recall aid use and cooperation. The IRQI is an ordinal index like the PCI, and a higher IRQI value indicates the respondent has higher reporting quality than another respondent with a lower IRQI value. We then constructed the summary Reporting Quality Index (RQI)² by subtracting the PCI from the IRQI.

² So that all components of the reporting quality index are consistent in having higher values indicate higher reporting quality, the equivalent of reverse coding the PCI is to subtract PCI from the IRQI.

The trends of these indices along the contact attempt continuum are charted in Figure 3. Due to increasingly smaller sample sizes, the indices become more volatile as the number of contact attempts increases. The RQI is also an ordinal index, where a higher RQI value indicates the respondent has higher reporting quality than another respondent with a lower RQI value.

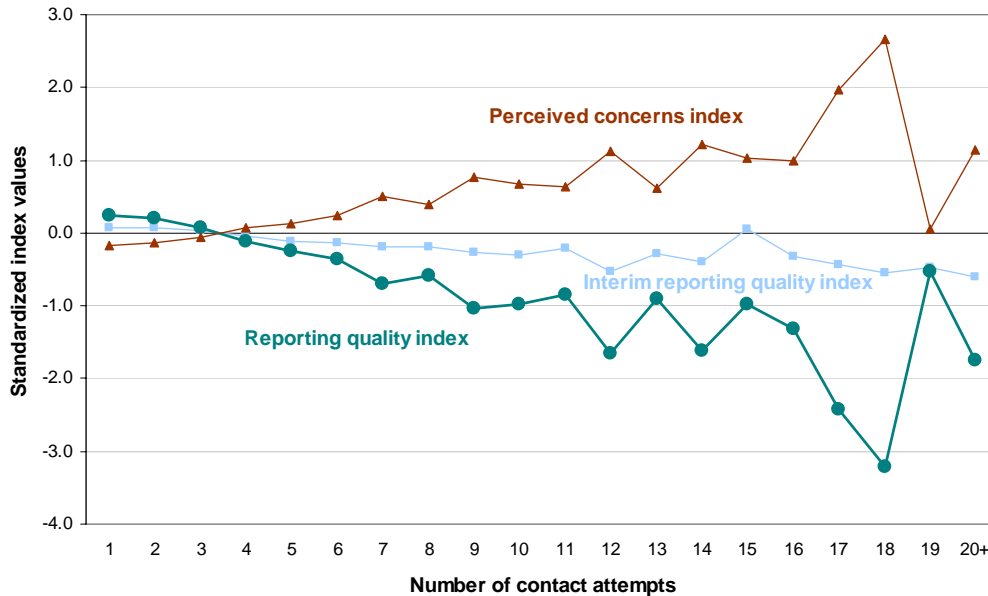


Figure 3: Reporting Quality Index by Number of Contact Attempts (Wave 1)

3. Findings

3.1 Reporting Quality

The reporting quality measures for the contact attempts groups are shown in Table 2. In all instances, interviews requiring eight or more attempts to complete had significantly more perceived concerns, higher rates of soft refusals, more “don’t know/refused” responses, and longer interviews. Interviews requiring eight or more attempts also had significantly lower rates of personal visit attempts, and less use of recall aids compared to interviews completed with less than five attempts. With the exception of longer interviews, the other measures indicate lower reporting quality for interviews completed in eight or more attempts compared to interviews completed with fewer attempts. Overall, the summary PCI and RQI indices exhibit a strong negative association with number of contact attempts.

Table 2: Reporting Quality Measures by Contact Attempt Group

	Number of Contact Attempts		
	1-4	5-7	8+
Number of Interviews	7,698	1,368	673
Reporting Quality Index ⁺	0.16	-0.38	-1.01 *
Perceived Concerns Index ⁺	-0.40	0.90	2.70 *
Intermediate Reporting Quality Index ⁺	0.05	-0.14	-0.28 *
Percent by Personal Visit	90.2	79.7	72.7 *
Percent Use Information Book	61.2	52.1	46.3 **
Percent Use Records	30.8	27.8	22.8 *
Percent of Expenditure Questions Answered with “Don’t Know/Refused”	4.5	5.3	6.6 **
Number Soft Refusals	2.3	11.1	19.8 *
Length of Interview (Minutes)	56.4	61.4	69.5 *

⁺ Standardized values.

* Group with eight or more attempts significantly different at p<0.5 from other groups.

** Group with eight or more attempts significantly different at p<0.5 from one to four attempt group.

3.2 Sample Composition

Compared to interviews completed with fewer than five attempts, respondents in the eight or more contact attempts group are younger, more likely to be black, Hispanic, college graduates, and residents of a Metropolitan Statistical Area (Table 3). They are more similar to respondents and CUs in completed interviews requiring five to seven attempts, a finding consistent with Curtin et al. (2000) who reported “much less respondent variation in third and later calls.”

Table 3: Sample Demographics⁺

	Number of Contact Attempts		
	1-4	5-7	8+
Number of Interviews	7,698	1,368	673
Respondent Characteristics			
Percent Male	43.6	43.5	41.3
Age	49.9	45.7	45.0
Percent Race*			
White	81.4	75.5	74.2
Black	10.0	15.0	14.8
Other	8.6	9.5	11.0
Percent Hispanic*	10.8	11.5	14.2

Table 3: Sample Demographics⁺

	Number of Contact Attempts		
	1-4	5-7	8+
Percent Education Attainment*			
Less than HS	14.8	10.8	13.1
HS Graduate	25.2	24.9	22.8
Some College	31.0	31.9	30.2
College Graduate	29.0	32.4	33.9
CU Characteristics			
Size of CU	2.43	2.46	2.42
Number of Persons < 18 Years	0.60	0.66	0.70
Number of Persons 65+ Years	0.35	0.20	0.18
Percent MSA*			
In MSA	85.1	87.3	90.2
Urban outside MSA	6.5	5.8	5.1
Rural outside MSA	8.4	6.9	4.8

⁺ Wave 1, unweighted.

* Group with eight or more attempts significantly different at $p < 0.5$ from group with one to four attempts.

3.3 Survey Estimates

A select number of expenditure estimates from unedited data is shown in Table 4. These expenditures were selected based on their relative frequency of reporting, and because they represent a range of expenditure values. Respondents requiring eight or more attempts to complete the first interview answered significantly fewer questions compared to respondents requiring one to four attempts (17.8 questions vs. 19.3). Among the selected expenditures we examined, there were no consistent trends across the groups. With the exception of Rent, Gasoline for Vehicles, and Subscriptions there were no significant differences among the groups in the other expenditures estimates examined.

Table 4: Selected Expenditure Estimates from Completed Interviews⁺

Variable or Expenditure Category	Number of Contact Attempts		
	1-4	5-7	8+
Number of Interviews	7,698	1,368	673
Number Expenditure Questions Answered*	19.3	18.4	17.8
Total Expenditures Reported	5,196	4,710	6,205
Rent*	199.5	248.1	245.4
Mortgage Payments	672.4	582.0	668.6
Phone Bill (1 month ago)	59.2	59.7	58.2
Internet (1 month ago)	4.2	4.3	4.2
Cable TV (1 month ago)	2.8	2.3	2.9
Utilities (1 month ago)	118.5	119.0	127.1
Major Appliances	24.4	23.9	25.5
Minor Appliances	117.4	101.0	110.4

Table 4: Selected Expenditure Estimates from Completed Interviews⁺

Variable or Expenditure Category	Number of Contact Attempts		
	1-4	5-7	8+
Clothing	132.6	140.1	170.1
Infants Clothing; Watches, Jewelry, & Hair Pieces	25.7	32.7	26.7
Clothing Services	2.0	1.4	1.5
Sewing Materials	2.4	2.5	1.7
Gasoline for Vehicles*	184.6	195.5	203.7
Oil for Vehicles	0.9	0.8	0.6
Current Month Subscriptions	12.6	18.9	12.4
Subscriptions*	45.9	49.3	31.4

⁺ Wave 1, unedited, 1 month recall.

* Group with eight or more attempts significantly different at $p < 0.5$ from group with one to four attempts.

3.4 Response Rates

While maximizing higher response rates is a major survey objective, it comes at a cost of greater resources spent on the collection effort, and possibly on data quality as well. While we do not have a comprehensive measure of data quality, our reporting quality index is a useful proxy for data quality. It allows us to graphically examine the tradeoff between response rates and data quality. The benefit of increasing response rates and the cost of declining reporting quality along the contact attempt continuum is shown in Figure 4. It appears that the threshold for the cost-benefit tradeoff of higher response rates and declining reporting quality occurs at approximately the eighth contact attempt and response rate of 65 percent.

3.5 Cost

While we do not show the analyses here, we estimate the cost savings from stopping at the seventh contact attempt to be about eight percent.

3.6 Summary of Tradeoffs

In summary, we examined the reporting quality, demographic characteristics, and expenditures levels of the contact attempt groups for Wave 1. We found that reporting quality was significantly lower in the eight or more contact attempts group compared to the one-to-four group, and demographically the respondents in this eight or more contact attempts group were younger, more likely black, Hispanic, college graduates, and residents of an MSA. There was no significant difference between the groups in total reported expenditures, although there was a general trend of higher reporting levels in the eight or more contact attempts group.

Considering the response rate-reporting quality tradeoff and the comparison of characteristics of the contact attempt groups previously described from the data used in this study, we propose the “optimal” threshold in the number of contact attempts to be *seven*.

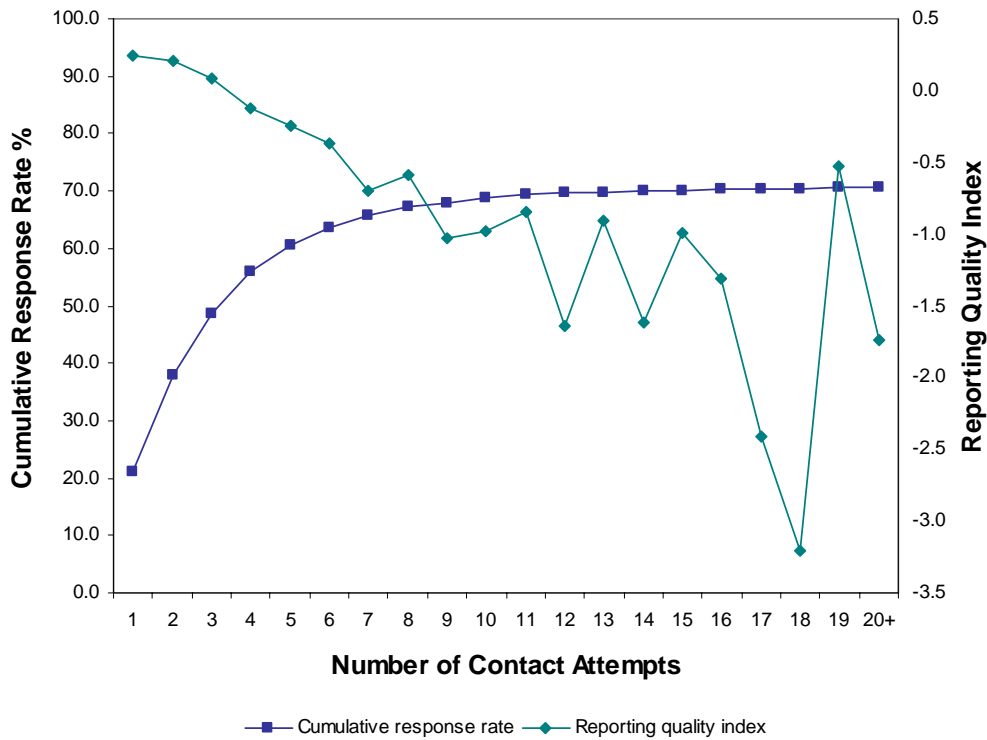


Figure 4: Cumulative Response Rate and Reporting Quality by Number of Contact Attempts

4. Evaluation of the Optimal Contact Attempt Threshold

4.1 Sensitivity Analysis

We examined the impact of excluding completed interviews requiring more than seven attempts on final edited expenditures for 13 major spending categories using Wave 2 data for the study period (see Table 5). These estimates were computed by reweighting the final calibration weights after excluding cases that required more than seven contact attempts to complete. Although the estimates from excluding interviews that required more than seven attempts were generally higher than the estimates including all attempts, there were no significant differences at the five percent level in expenditure estimates between these two groups. The magnitude of the standard errors for these estimates was also similar, with the exception of Personal Insurance.

Table 5: Comparison of Summary Expenditure Estimates Excluding Interviews Requiring More than Seven Attempts⁺

Expenditure Category	1 to 7 Attempts (n=12,418)		All Attempts (n=13,742)		Difference (1-7 attempts group) – (All attempts group)	
	\$	SE*	\$	SE	Diff \$	SE (diff)
Total Expenditures	11,907	136.0	11,596	133.7	311	190.7
Housing	4,058	63.5	3,973	61.0	85	88.0

Table 5: Comparison of Summary Expenditure Estimates Excluding Interviews Requiring More than Seven Attempts⁺

Expenditure Category	1 to 7 Attempts (n=12,418)		All Attempts (n=13,742)		Difference (1-7 attempts group) – (All attempts group)	
	\$	SE*	\$	SE	Diff \$	SE (diff)
Transportation	2,217	37.7	2,139	44.9	78	58.6
Food	1,678	13.1	1,650	13.7	28	18.9
Apparel	341	8.1	341	10.1	1	12.9
Health	680	9.7	661	11.6	19	15.1
Education	249	14.2	240	15.3	9	20.9
Entertainment	602	14.4	587	13.1	15	19.4
Personal Care	76	1.6	74	1.2	2	2.0
Personal Insurance	1,167	25.0	1,125	16.9	42	30.1
Reading	31	0.5	30	0.7	1	0.9
Alcoholic Beverages	89	2.6	90	2.8	-1	3.8
Tobacco	81	2.1	84	2.5	-3	3.3
Cash Contributions	479	26.6	454	21.8	25	34.4

⁺ Wave 2 data, with final calibration weights.

* SEs computed using BRR weights in SUDAAN to account for the CE's complex survey design.

4.2 Estimates of Relative Nonresponse Bias

Our examination of the demographic characteristics indicated differences between respondents in the eight or more attempts group relative to respondents in the one to four attempts group. Omitting the former group from computation of survey estimates may contribute to nonresponse bias. We estimate the relative nonresponse bias for the key aggregate expenditures estimates following a similar methodology to those used in recent studies of potential nonresponse bias for the CE (King et al. 2009). One of those studies used “harder-to-contact” respondents as proxy nonrespondents, where a respondent was classified as a proxy nonrespondent if more than 45 percent of attempts to contact the respondent resulted in noncontacts; this threshold was chosen because it yielded a response rate that was similar to the CE's actual response rate during the study period. The current paper study offers an alternative measure of proxy nonrespondents, and defines them as respondents requiring eight or more contact attempts to complete an interview.

$$\text{Relative Bias } RB(\hat{Z}_j) = \frac{\text{Bias}(\hat{Z}_{jR})}{\hat{Z}_{jT}} = \frac{\hat{Z}_{R,j} - \hat{Z}_{T,j}}{\hat{Z}_{T,j}} = \frac{n_N}{n_T} (\hat{Z}_{jR} - \hat{Z}_{jN})$$

where:

- \hat{Z}_{jT} = mean expenditure estimate for expenditure category j from the total sample (which is treated as the “truth”)
- \hat{Z}_{jR} = mean expenditure estimate for expenditure category j from respondents
- \hat{Z}_{jN} = mean expenditure estimate for expenditure category j from proxy nonrespondents
- n_N = number of proxy nonrespondents
- n_T = total sample

A negative relative bias for an expenditure estimate indicates that omitting the eight or more contact attempt cases potentially results in an underestimation of expenditures, whereas a positive relative bias indicates an overestimation of the expenditure. The *magnitude* of the relative bias indicates the extent to which the “true” value is changed by omitting nonrespondents in the computation of mean expenditure estimates. Where the 95 percent confidence interval of the relative bias includes zero, the relative bias probably reflects sampling variability and the expenditure estimate may not be subject to nonresponse bias. The underlying assumption behind these computations is that nonresponse is the only source of bias.

The estimates of relative nonresponse bias are shown in Table 6. While total expenditures is not subject to nonresponse bias, the inclusion of the eight-or-more attempts group respondents *lowered* the mean expenditure estimates in several categories (Health, Cash Contributions, and Reading) but *raised* the mean expenditure estimates in others (Housing, Apparel and Services, and Personal Care). The magnitude of relative bias estimate was largest in Apparel and Services, Health, and Cash Contributions, although these categories accounted for relatively small shares in total spending (3.0 percent, 5.9 percent, and 3.0 percent, respectively).

Table 6: Base-weighted Summary Expenditures and Estimates of Relative Bias⁺

Expenditure Category	% Share of Total Exp.	Mean Expenditure Estimates \$			Relative Bias %		
		Proxy NRs [^] (8+ attempts)	Rs [^] (1 to 7 attempts)	All Rs	Estimate	95 LCI	95 UCI
Number of Interviews		1,324	12,418	13,742			
Total Expenditures	100.0	12,082	11,656	11,697	-0.35	-1.20	0.40
Housing	33.8	4,371	3,914	3,958	-1.11 *	-2.02	-0.27
Transportation	18.2	2,099	2,132	2,129	0.15	-1.31	1.27

Table 6: Base-weighted Summary Expenditures and Estimates of Relative Bias[†]

Expenditure Category	% Share of Total Exp.	Mean Expenditure Estimates \$			Relative Bias %		
		Proxy NRs [^] (8+ attempts)	Rs [^] (1 to 7 attempts)	All Rs	Estimate	95 LCI	95 UCI
Food	14.6	1,782	1,696	1,705	-0.49	-1.08	0.13
Personal Insurance	9.6	1,200	1,119	1,127	-0.69	-1.97	0.52
Health	5.9	559	699	685	1.96*	1.15	2.83
Entertainment	5.2	598	604	603	0.09	-0.84	1.01
Cash Contributions	4.0	367	476	465	2.25*	0.35	3.56
Apparel	3.0	426	341	349	-2.34*	-3.79	-0.87
Education	2.2	247	259	258	0.47	-4.20	4.96
Alcoholic Beverages	0.7	99	85	86	-1.59	-3.30	0.27
Tobacco	0.7	83	82	82	-0.09	-2.37	1.90
Personal Care	0.6	83	73	74	-1.30*	-2.20	-0.38
Reading	0.3	26	31	31	1.72*	0.07	3.11

[†] Data are from Wave 2, April 2006 through March 2008 (N=13,742). Proxy nonrespondents are those in CUs who require eight or more contact attempts, comprising 9.6 percent of N.

[^] NRs: nonrespondents; Rs: respondents.

* Significantly different from 0 at p<0.05.

5. Discussion

5.1 Summary

In summary, we examined the reporting quality, demographic characteristics, and expenditures levels of the contact attempt groups for CEQ data. We found that reporting quality was significantly lower in the eight-or-more attempts group compared to the one-to-four group. Demographically, the respondents in the eight-or-more group were younger, and more likely to be black, Hispanic, college graduates, and residents in an MSA than respondents in the one-to-four attempts group, but more similar to respondents in the five-to-seven attempts group. There was no significant difference between the groups in total reported expenditures, although there was a general trend of higher reporting levels in the eight-or-more group.

It also appears that the “cost-benefit” tradeoff in terms of response rates and reporting quality with additional contact attempts occurs at about the eighth contact attempt, where the response rate is 65 percent. From this observation, and the consideration of the group profiles, we proposed the “optimal” threshold in the number of contact attempts to be seven. Given the cost assumptions we made, this would result in a cost savings of about 8 percent. In addition, stopping at the seventh contact attempt appeared to have negligible effect on key estimates, and did not have an adverse effect on estimated relative nonresponse bias.

5.2 Limitations

As several of the primary analytical methods used in our study are based on descriptive procedures (principal components analysis and cluster analysis), our recommendation should be considered from an exploratory point of view rather than an inferential one.

There were several limitations with the data we used:

- a. Our analyses are based on historical data. As Curtin et al. (2000) point out, truncating a sample to yield a lower response rate is not the same as conducting a survey designed to yield a lower response rate. If only a limited number of contact attempts were permitted for a survey, the attempts would probably be scheduled differently – for example, with a more focused effort to attempt contact at productive times. At the same time, from a forward-looking standpoint, telling interviewers that they can stop at eight attempts might have unforeseen consequences, and lead to a general degradation of quality, or at least to a slippery slope of worsening performance.
- b. We used Wave 1 data to examine cluster profiles. Wave 1 data do not undergo the internal edit processes that check for consistency and outliers. Therefore, data inconsistencies and errors that may occur during the survey data collection process were not captured and adjusted by the survey's routine data editing processes.
- c. Each contact attempt is based on the interviewer's self-report. There is currently no mechanism to ensure that each contact attempt is recorded in the Contact History Instrument, so that the number of contact attempts for each case may not be accurate.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. The authors thank Jeffrey M. Gonzalez, Manjusha Kalavapudi, Christopher Laskey, and Shiao-Lin Shirley Tsai for their very helpful contributions to this study.

References

- Bates, N. and Creighton, K. (2000). The last 5 percent: what can we learn from difficult/late interviews? *Proceedings of the Section on Government Statistics and Section on Social Statistics*, American Statistical Association, pp. 120-125, Alexandria: VA.
- Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, Vol. 64 No. 4, pp. 413-428.
- Dennis, J.M., Mathiowetz, N.A., Saulsberry, C., Frankel, M., Srinath, K.P., Roden, A., Smith, P.J., and Wright, R.A. (1999). Analysis by number of call attempts. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Pete Beach, FL.
- Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 646-675.

- Heerwegh, D., Abts, K., and Loosveldt, G. (2007). Minimizing survey refusal and noncontact rates: do our efforts pay off? *Survey Research Methods*, Vol. 1 No. 1, pp. 3-10.
- Keeter, S., Miller, C., Kohut, A., Groves, R. and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey, *Public Opinion Quarterly*, Vol. 64, 2, p. 125-148.
- King, S.L., Chopova, B., Edgar, J., Gonzalez, J.M., McGrath, D., and Tan, L. (2009). Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey. Paper presented at the 2009 Joint Statistical Meetings of the American Statistical Association, Washington, DC.
- Safir, A., and Goldenberg, K. (2008). "Mode Effects in the U.S. Consumer Expenditure Quarterly Interview Survey." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, New Orleans, LA.
- Srinath, K.P., Giambo, P. and Arday, S. (2003). An Application of a Method of Determining Optimum Number of Call Attempts in a Telephone Survey. *Proceedings of the Section on Survey Methods Research*, American Statistical Association, pp.4022-4027, Alexandria: VA.
- Tan, L. and Tsai, S. (2008). An exploration of respondent concerns perceived by the interviewer for the Consumer Expenditure Interview Survey. Internal BLS-DCES Internal Report.