

Towards Determining an Optimal Contact Attempt Threshold for a Large-Scale Personal Visit Survey⁺

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

While improving the design of the survey instrument and increasing the options for more convenient and/or efficient data capture are two important ways for maximizing returns on the survey operations budget, there is an additional dimension in personal-visit surveys for discretionary action that can make a difference – setting a maximum threshold for the number of contact attempts, or the level of effort, expended to resolve a case. In particular, the cost of contact attempts can be significant for personal-visit surveys with eligible sample units spread over an expansive geographic area. Before fielding a large-sample test of a proposed “optimal” contact attempt threshold, the authors initiated this study as a verification of findings from an earlier study (Safir and Tan 2009) that recommended 7 attempts as the threshold for resolving a sample unit’s final disposition. Using more recent data and additional evaluation metrics, the authors perform a retrospective analysis of first wave panel data from the Consumer Expenditure Interview Survey (CEQ) collected between April 2012 and March 2014.

Keywords: Personal-visit survey, reporting behavior, contact history, field operations

1. Introduction

As with other federal statistical agencies, the Bureau of Labor Statistics continues to operate under budget constraints. While improving the design of the survey instrument and increasing the options for more efficient data capture are two important ways for maximizing returns on the survey operations budget, there is an additional dimension in personal-visit surveys for discretionary action that can make a difference – setting a threshold for the number of contact attempts, or the level of effort, expended to resolve a case. In particular, the cost of contact attempts can be significant for personal-visit surveys with eligible sample units spread over an expansive geographic area.

In a retrospective analyses of Wave 1 data from the Consumer Expenditure Interview Survey (CEQ) collected between April 2006 and March 2008, Safir and Tan (2009; henceforth referred to as the *2009 Study*) recommended the threshold for resolving a sample unit’s final disposition be 7 attempts. Before fielding a large-sample experiment to test this recommendation, it was evaluated against more recent data.

Similar to the *2009 Study* where response rates and indicators of reporting behavior quality were the primary evaluation metrics, this current retrospective study expanded on the set of evaluation metrics to include a wider variety indicators of reporting behavior, including representativity “R”-indicators (Schouten, Cobben, and Bethlehem 2009) to evaluate the composition of response on a limited set of sample unit

⁺The opinions expressed in this paper are those of the authors and do not represent official policy of the Bureau of Labor Statistics. The authors thank Brett McBride and Jeffrey M. Gonzalez for their review and feedback.

characteristics to examine the impact of the recommended contact attempt cutoff on sample representativeness with respect to those characteristics.

1.1 The Study Sample

The data was based on the Consumer Expenditure Interview Survey (CEQ), a national household survey on spending. The CEQ has a rotating panel design, and is a computer assisted personal interview (CAPI) survey, although telephone interviews do occur.¹ 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.

This study uses CEQ and CHI data for Wave 1 sample units spanning April 2012 through March 2014. The final dispositions of the sample units in the *2009 Study* and the current study are shown in Table 1 - the rate of completed interviews had declined from 59.1 percent to 54.6 percent, and the refusal rate increased from 11.2 percent to 13.8 percent. In the *2009 study*, 86 percent of all cases and 89 percent of completed interviews were resolved by the 7th attempt; in the current study, the resolution rates fell to 78 percent of all cases and 84 percent of interviews. When the study sample was restricted to eligible sample units with at least 1 contact made (as recorded in the CHI; contact with the sample unit is needed to obtain information on interviewers' observations about sample units' reactions to the survey request, i.e. the doorstep concerns), the cumulative resolution rate of eligible cases was 76 percent (of N=18,031) by the recommended 7th contact attempt (Table 2); the proportion of completed interviews resolved at this cut-off remained at 84 percent. Findings presented in this report are based on this restricted sample of eligible units with at least 1 contact recorded in their contact attempt history. The overall response rate was 75.2 percent (of N=18,031). All analyses were run on unedited data.

Distribution of contact attempts. There was a total of 100,775 contact attempts recorded in the CHI for the study sample (N=18,031, mean of 5.6 contact attempts), of which 49,497 contact attempts were attributed to sample units requiring 7 or fewer contact attempts to resolve (n=13,631, mean of 3.6 contact attempts), and 51,278 attempts were attributed to sample units requiring 8 or more attempts to resolve (N=4,400, mean of 11.7 attempts). The overall distribution of contact attempts to resolve cases was highly skewed to the right, with the median number of attempts at 4, and the mode at 2 (Figure 1a). For completed interviews, the mode was 2 and the median 4 attempts (compared to 3 in the *2009 Study*); for unit nonresponse, the mode was 6 and median 7 attempts (Figure 1b). By personal visit (visit attempts) for interviews, both the mode and median were 2 visit attempts; for unit nonresponse, the mode was 3 and median 4 visit attempts.

2. Methods

Comparison groups. Since we were testing the *2009 Study's* recommendation of 7 contact attempts, the two comparison groups were sample units who required more than 7

¹ For more details on the Consumer Expenditure Survey Program, see <http://www.bls.gov/cex>

contact attempts to resolve their cases (referred to as the “*high attempt group*”, N=13,631) and those requiring 7 or fewer attempts (the “*other group*”, N=4,400).

The demographic characteristics of sample units who completed the survey request by these comparison groups are shown in Table 3.

Indicators of sample units’ observed reluctance (“doorstep concerns”, Bates et al. 2008) from their contact attempt history

Indicators for sample units’ reluctance at the survey request were created based on information recorded by interviewers in CHI. In addition to various characteristics of contact attempts (number, mode of attempt, duration between first and final attempt, soft refusals (contact made but no interview because the sample unit member was reluctant), we also created an indicator to flag if there was at least one change in interviewer in the sample unit’s contact attempt history.

Doorstep concerns. Prior research (Kopp et al. 2013, McBride and Tan 2014, and Safir and Tan 2014) had found that interviewer observations about contacted sample units’ initial reactions to the survey request in the form of “doorstep concerns (DS) themes” recorded in the CHI was associated with survey response propensity in the CEQ. While the 2009 Study had incorporated doorstep concerns into the analysis as an ordinal composite score to rank the sample units according to the level of their concerns, the current study incorporated specific doorstep concern themes instead of a summary score in analyses. These themes are based on “ad hoc” groupings of DS items as in Kopp et al. (2013) and shown in Table 4 – this approach has the advantage of being consistent in groupings of items and reasonable in interpretation. If a sample unit’s contact attempt history indicated that a DS item was reported at least once, the DS theme indicator corresponding to that DS item is flagged for the unit. An indicator was also created if no doorstep concerns was ever reported in the sample unit’s contact attempt history.

Response composition.

The primary motivation behind striving for higher response rates is to minimize the potential for nonresponse bias arising from differences between respondents and nonrespondents. However, recruiting more respondents to participate in the survey need not translate to more representative response.

Subgroup response rates. Examining variation in subgroup response rates provide insights to response composition. Among the demographic characteristics associated with expenditures and for which there are no missing values for the entire sample are household size, homeownership status, and urbanicity of the dwelling unit. The subgroup response rates for these characteristics, units grouped by the recommended contact attempt cutoff, and units grouped by any doorstep concerns are shown in Figure 2.

R-indicators. R-indicators provide a way to evaluate the composition of response with respect to a select set of characteristics of interest that are known for both respondents and nonrespondents, and which are observed external to the survey (Schouten, Cobben, and Bethlehem 2009). The R-indicator measures the “distance” of the individual sample units’ response propensities from the overall average response propensity of the subsample; it ranges from 0 to 1. The larger the variability (lower R-indicator value), the higher the potential for selective response and the violation of Missing-At-Random with respect to the predictors in the response propensity model, and thus higher potential for nonresponse bias.

The basic input to computing R-indicators are sample units' probability of responding to the survey. We estimated response propensities for the study sample using logistic regression, with two different sets of predictors:

- The 1st model (Model DMG) included indicators for single member CU (for household size) , dwelling structure located within an MSA, and homeownership status
- The 2nd model (Model DSDMG) predictors comprised the 3 predictors in Model DMG, plus an indicator for whether any doorstep concerns were observed.

For each model, the R-indicator was computed for subsets of the study sample upon the completion of select number of contact attempts (1 3, 5, 7, 8, 12) as well as for the entire study sample.

The *sample based R-indicator* was computed as:

$$R = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^N d_i (\hat{\rho}_i - \hat{\rho})^2}$$

where i is the index for the sample unit, and d_i is the design weight for unit i , and the weighted sample average of response propensities: $\hat{\rho} = \frac{1}{N} \sum_{i=1}^N d_i \hat{\rho}_i$. (Cobben 2009; Heif, Schouten, and Shlomo 2014).

The bootstrap 95 percent confidence interval for the R-indicator was computed using 500 bootstrap samples, sampling with replacement using the same sample size at the selected number of contact attempts.² At each of the select contact attempt, the 2.5th and 97.5th percentiles from the 500 bootstrap R-indicators were used to estimate the bounds of the confidence interval.

Indicators of reporting quality

Without direct measures of reporting quality, we examined indicators of reporting behavior that could be reasonably interpreted as supportive of good reporting quality. Some of these indicators are based on interviewer assessments reported upon completion of the interview, such as the mode of the interview and the use of recall aids; other indicators are based on respondents' reports - don't know/refused" answers to expenditure questions, total reported expenditures, endorsement of filter questions to selected expenditure categories, and reporting of "combined items" reporting.

Endorsement of filter questions for selected expenditure categories and reporting of "combined items." Reluctant respondents may be more likely to avoid reporting purchases, or less likely to itemize their purchases in detail ("combined items" reporting), to complete the survey more quickly. The former would result in under-reporting due to omission, and the latter require data editing to allocate those expenditures reported in aggregate among target items. We examined the endorsement rate of filter questions for selected expenditure categories and the prevalence of reporting of "combined items" within a category (conditioned on endorsement of that category's filter question) as indicators of reporting behavior quality. The subset of expenditure categories that were

² Cobben (2009, page 134) used 200 bootstrap samples for all studies and found that the estimate of the standard error had converged for bootstrap samples of size less than 200.

selected for analysis were those that we thought could reasonably be expected of every household to incur, and whose locations in the survey were fairly spread out. For each expenditure category, two indicator variables were created. The first was a category endorsement indicator which was flagged as 1 if the respondent reported “Yes”, or if the respondent reported at least 1 item to the filter question (depending on the format of the filter question). The second was a category ‘combined items’ indicator which was flagged as 1 if among the items reported in response to the filter question, there was at least 1 item code that represented “combined items”. For each category, the *category endorsement rate* was computed as the proportion of the study sample with the endorsement indicator equal to 1. The *category combined item reporting rate* was computed as the proportion of the study sample who endorsed the filter question with the combined items indicator equal to 1.

3. Findings

3.1 Characteristics of contact attempts and doorstep concerns of respondents.

Compared with other respondents, high attempt respondents required 3 times more attempts (overall and visit attempts) to resolve their interviews, more contacts needed to complete interviews, twice more likely to result in soft refusals after contact, their case remain unresolved twice as long, and 4 times more likely to involve at least one change in interviewer (Table 5). The high attempt respondents were also less likely to be perceived by interviewers to have doorstep concerns –27 percent of them perceived to have no doorstep concerns compared to 61 percent of the other respondents. Among DS themes, “Time” was most prevalent for both groups; however, with the exception of “Survey content/Privacy”, the prevalence of all other DS themes was more than twice higher among the high attempt group.

These indicators together suggest that high attempt respondents were perceived to exhibit a greater degree of pre-survey reluctance, and they required the engagement of more field resources and time to resolve their interviews compared with other respondents.

3.2. Subgroup response rates. The overall response rate at the completion of all contact attempts was 75.2 percent. Comparing the variation in response rates among subgroups of each of the 5 characteristics, the largest range of variation in subgroup response rates was by doorstep concerns (units without doorstep concerns were almost two times more likely to participate in the survey) and the smallest variation among the three demographic characteristics (within 9 percentage points) (Figure 2).

3.3. R-indicators. An examination of the subgroup response rates in Figure 2 had indicated that variability in response rates between categories of a subgroup depended on the characteristic of interest. Specifically, the variability in subgroup categories for household size, urbanicity, and homeownership status was relatively small compared with the variability in subgroup categories by contact attempt characteristics. The R-indicator for models DMG and DMDGS and cumulative response rate attained upon completion of select contact attempts are shown in Table 6. Upon completion of contact attempt number 1, 7, and all contact attempts, the cumulative response rates were 10.1 percent, 63.1 percent, and 75.2 percent, respectively; the R-indicator for Model 1 (DMG) were 0.99, 0.93, and 0.93, respectively, indicating that consistently low variability of response probabilities is expected with respect to household size, urbanicity, and homeownership status with additional contact attempts. In contrast, the R-indicator for

Model 2 (DSDMG) was 0.94, 0.70, and 0.64, indicating increasing variability in response probabilities with respect to household size, urbanicity, and homeownership status, and not having doorstep concerns.

3.4 Indicators of Reporting Quality

Duration. While the difference in the median duration to complete the interview was about 6 minutes longer for the high attempt group (whose median duration was 59.4 minutes), the duration on the expenditure sections (1 through 20) of the survey was similar between the two respondent groups (median of 36 minutes, see Table 7).

Mode. The prevalence of completing the interview entirely by personal visit (which is preferred by design) was lower among the high attempt group (64 percent, compared with 85 percent among other respondents).

Use of recall aids. The prevalence in use of both types of recall aids (records and information book) for more than 50 percent of the interview was less than one-fifth of respondents in both groups but it was 5 percentage points lower for the high attempt group; the high attempt group was also higher in prevalence of “less than 50 percent or no use of both types for recall aids (56 percent compared with 47 percent among other respondents).

Endorsement of filter questions and combined items reporting. High attempt respondents generally endorsed the filter questions for the expenditure categories examined at similar or lower rates as other respondents, with difference of about 10 percentage points lower in the categories of “*Medical services: payments*” (48 percent vs. 39 percent, respectively) and “*Miscellaneous*” (53 percent vs. 46 percent).

Conditioned on endorsing the filter question of a category, the higher contact attempt group reported “combined items” at similar or higher rates compared with other respondents (see Table 7). The largest difference of 10 percentage points between the two groups was in “Apparel services” (52 percent vs 62 percent), this may largely be due to the small number of endorsers in both groups (less than 4 percent) for this category.

Multivariate regressions

a. Expenditure reporting. To examine the association between reported total expenditures (unedited) with the number of contact attempts, we estimated a logistic regression model with the dependent variable being an indicator of whether a sample unit’s reported total expenditures exceeded the median value, control variables to be indicators for homeownership status, single household member (household size), use of recall aids more than 50 percent of the time, and survey length (for sections 1 through 20) greater than the median as control variables, and the predictors of interest to be the number of contact attempts and doorstep concern themes. There was no difference between high attempt and other respondents in their odds of reporting total expenditures greater than the median value; however, the individual doorstep concern themes varied in significance and effect (Table 8).

b. Endorsement of filter questions. To examine the association of the number of filter questions endorsed with contact attempts, the same set of predictors used in an ordinary least squares regression model. The number of filter questions endorsed was negatively associated with more than 7 contact attempts and more than 3 visit attempts, and the individual doorstep concern themes varied in significance and effect (Table 9).

While the length of time to complete the interview and overall expenditures reported per se are not definitive about the quality of reporting behavior, the trends of lower prevalence in the use of recall aids more than 50 percent of the time, lower prevalence in the conduct of the interview entirely by personal visit, higher proportion of “don’t know/refused” expenditure reports, similar or lower endorsement rate of filter questions to selected expenditure categories, and similar or higher rate of “combined items” reporting among the high attempt respondents together suggest at best no improvement in the quality of overall reporting behavior among high attempt respondents compared with the other respondents with respect to the metrics examined. The multivariate regressions had highlighted the negative association of doorstep concern themes “Not interested/Hostility” and “Survey content / Privacy” with reporting quality in terms of overall expenditure reporting and endorsement of filter questions, and both these themes have a higher prevalence among high attempt respondents (15.5 percent and 29.9 percent, respectively among high attempt respondents compared with 5.8 percent and 17.6 percent, respectively among other respondents, see Table 5.)

4. Summary & Discussion

The findings from this analysis support the 2009 Study’s recommendation of 7 contact attempts as the threshold. Cost were substantially higher for the high attempt group: high attempt respondents required the engagement of more field effort in terms of personal visit contact attempts, contacts, and multiple interviewers to resolve their interviews compared with other respondents. From a total survey error approach, and with respect to the trade-off between the cost of procuring higher survey response rates and the benefit of significant data collection cost savings, without evidence that the higher response rates are associated with decreasing the potential for nonresponse bias and decreasing measurement error (as assessed by reporting quality indicators), the authors recommend that CE field test a contact attempt threshold of 7.

Limitations. Due to the criterion that only sample units with whom interviewers recorded at least one contact in the CHI was included in the study sample, in theory, the nonrespondents in this study were restricted to nonresponse due to refusals.³

We lacked direct measures of reporting quality. Ideally, the use of recall aids, especially records, would be a very useful indicator. However, the extent of the usage of records and information booklet questions are asked of the interviewers at the end of the survey and based solely on the interviewer’s assessment and recall.

With the paucity of socio-demographic information on the sample frame, we relied on the CHI for information that would be available for both respondents and nonrespondents related to their contact attempt characteristics and perceived sample units’ pre-survey doorstep concerns. Among the evaluation metrics we used, these characteristics from the CHI were generally found to have higher discrimination power

³ However, less than 3.5 percent of the sample units in the study sample were found to be coded as nonresponse due to “no one home” or “temporary absent, see Table 3.

on unit response and quality of reporting behavior. However, the CHI relies solely on the interviewer recording every contact attempt entry and his / her subjective assessment of the contacted sample unit's reaction to the survey request and when that information is recorded. In addition, the meaning of the doorstep concern themes used in this analyses have not been formally tested cognitively or otherwise for their practical meaning. A better understanding of the correct interpretation of these themes is necessary to inform the design of interventions that would be based on them. Nonetheless, these doorstep concern themes have consistently demonstrated differential qualitative effects on survey response in previously cited recent CEQ studies.

References

- Bates, N., J. Dahlhammer, and E. Singer (2008). Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse. *Journal of Official Statistics*, Vol.24, No.4, 2008. pp. 591–612.
- Cobben, F. (2009). Nonresponse in Sample Surveys: Methods for Analysis and Adjustment. Statistics Netherland.
<http://www.cbs.nl/NR/rdonlyres/2C300D9D-C65D-4B44-B7F3-377BB6CEA066/0/2009x11cobben.pdf>
- Kopp, B., McBride, B., and Tan, L. (2013). An exploratory study on the association of doorstep concerns with three survey quality measures for the CE Interview Survey. Division of Consumer Expenditure Survey Internal Report, Bureau of Labor Statistics.
- McBride, B., and Tan, L. (2014). Quantifying CHI doorstep concerns as risk factors of Wave 1 nonresponse for the CE Interview Survey. Division of Consumer Expenditure Survey Internal Report, Bureau of Labor Statistics.
- Safir, A. and Tan, L. (2009). Using Contact Attempt History Data to Determine the Optimal Number of Contact Attempts: an Exploratory Study. Bureau of Labor Statistics, DCES Internal Report.
- Safir, A. and Tan, L. (2014). Estimating the effect of initial expressions of reluctance and the risk of first wave nonresponse in the Consumer Expenditure Interview Survey. Bureau of Labor Statistics, DCES Internal Report.
- Schouten, B., Cobben, F., and Behrlehem, J. (2009). Indicators of the representativeness of survey response. *Survey Methodology*, 35(1), 101-113.

Table 1. Wave 1 frame sample (includes ineligible units, with at least 1 contact in the CHI)

	INSCOPE CASES = INTERVIEWS + Type A NONRESPONSE				OUT-OF-SCOPE Type B & C	Row Total
	Type A nonresponse					
	Interviews	Refused	Noncontact	Other		
2009 Study: April 2006-March 2008						
N	12,844	2,433	1,080	540	4,836	21,733
% of N	59.1	11.2	5.0	2.5	22.3	100.0
Current study: April 2012-March 2014						
N	13,655	3,454	1,320	1,136	5,432	24,997
% of N	54.6	13.8	5.3	4.5	21.7	100.0

Table 2. Cumulative resolution rates by final disposition categories for current study sample (with at least 1 contact reported in contact attempt history)

	All	Interviews	Type A refused	Type A language	Type A other	Type A no one home	Type A temp absent	
N	18,301	13,567	3,071	31	790	462	110	
% of N	100.0	75.2	17.0	0.2	4.4	2.6	0.6	
No. attempts to resolution	Column percent distribution							Row n
1	10.32	13.5	0.9	0.0	0.4	0.0	0.9	1,860
2	25.0	31.9	4.8	9.7	3.4	0.9	3.6	2,647
3	38.9	48.0	13.3	19.4	7.8	3.5	7.2	2,498
4 (median)	50.8	61.0	22.7	35.5	14.5	10.0	17.2	2,147
5	60.7	70.9	33.7	58.1	24.4	15.8	23.6	1,797
6	69.0	78.5	44.4	67.8	33.5	26.4	37.2	1,498
7 (2009 Study)	75.6	83.9	54.9	67.8	43.5	35.7	43.6	1,184
8	80.5	87.9	62.6	67.8	50.5	45.0	52.7	888
9	84.8	91.0	70.1	71.0	58.6	55.0	60.0	770
10	88.1	93.3	76.2	74.2	65.8	61.5	71.8	597
11	90.7	95.1	81.1	80.7	71.4	66.0	79.1	467
12	92.8	96.4	85.4	83.9	76.8	70.8	83.6	380
13	94.4	97.3	88.4	93.6	81.7	75.1	86.3	282
14	95.7	97.9	91.3	96.8	84.6	80.7	89.0	230
15	96.6	98.4	93.3	96.8	87.5	85.2	90.8	179
16	97.4	98.8	95.0	100.0	89.8	89.1	90.8	144
17	97.9	99.0	95.7	100.0	91.8	92.1	92.6	81
18	98.3	99.2	96.8	100.0	93.2	92.7	92.6	82
19	98.7	99.4	97.4	100.0	94.2	94.9	94.4	61
20+	100.0	100.0	100.1	100.0	99.9	99.9	99.9	239
								18,301

Table 3 Demographic characteristics of sample units completing interviews in Wave 1

	N	No. contact attempts	
		1-7	8+
Sample size (N)		11,370	2,197
<i>Reference person</i>			
Age (years)	*	51.3	46.2
Age group distribution(%N)	*		
lte 29		13.0	16.4
>29-45		26.4	34.3
>45-64		35.6	35.8
gte 65		25.0	13.4
Race (%N)	*		
White		81.4	77.1
Black		11.3	14.4
Other		7.2	8.5
Hispanic origin (%N)	*	12.5	16.4
Male (%N)		46.9	45.0
Education attainment (%N)			
< High School	[*PV]	12.7	12.0
High School graduate		24.7	23.4
Some college		31.1	31.6
College graduate		31.5	33.1
<i>CU characteristics</i>			
No. members in CU (%N)	*		
1		29.4	29.4
2		33.1	29.0
3		15.1	16.5
4+		22.4	25.1
No. members aged 18 and under	*	0.58	0.71
No. members aged 65 and older	*	0.39	0.21
No. earners	*	0.76	0.82
Dwelling in MSA(%N)	*	86.6	89.7
Homeowner(%N)	*	64.3	59.0

Table 4. Grouping of doorstep concern items to form doorstep concern themes (based on Kopp et al. 2013)

Doorstep concern theme	Prior Wave and Other	Time	Not interested/ Hostility	Survey Content/ Privacy	Gatekeeping
Theme Indicator variable	iothpw	itime	inh	icp	igate
CHI doorstep concern items	16 - R. requests same FR 17-info previously given 18 - too many questions previously 19- too many interviews 20 - intv too long previously 21-intends to quit survey 23-other	2-too busy 3-intv too time consuming 4-brk appt-puts off FR indefinitely 5-scheduling difficulty	1-not interested 11-hangs up/slams door 12-hostile /threatening	6-survey voluntary 7-privacy concerns 8-anti-govt 9-does not understand survey 10-survey content not applicable	13-othr hh members say don't do survey 14-talk to specific hh member 15-family issues
No. items:	6	4	3	5	3

Table 4. Grouping of doorstep concern items to form doorstep concern themes (based on Kopp et al. 2013)

Doorstep concern theme	Prior Wave and Other	Time	Not interested/ Hostility	Survey Content/ Privacy	Gatekeeping
Theme Indicator variable	iothpw	itime	inh	icp	igate
Prevalence (% of N=18,031; >1 theme may be observed for a CU)	iothpw: 17.7 ioth: 13.5; ipwave: 5.5	33.5	19.1	27.5	7.0

Table 5. Contact attempt and doorstep concern characteristics of sample units completing interviews in Wave 1

		No. contact attempts	
		1-7	8+
Sample size (N)	N	11,370	2,197
<i>Contact attempt characteristics</i>			
All contact attempts			
Number	Mean	3.4	11.1
	Median	3	10
Resulting in contact (% attempts)	Mean	59.4	27.8
Duration between 1 st and final attempt (days)	Mean	8.9	20.5
Visit attempts			
Number	Mean	2.4	6.7
	Median	2	6
Proportion of contact attempts	Mean	0.77	0.61
	Median	0.80	0.63
Contact (with sample unit member)			
Number	Mean	1.68	2.9
	Median	1	3
Resulting in soft refusals (% Contacts)	Mean	1.9	5.4
Ever changed interviewer within a wave (%N)	Mean	7.9	37.6
<i>Doorstep concerns perceived (CHI)</i>			
No doorstep concerns (%N)	Mean	60.7	27.1
Doorstep concern Themes			
Number	Mean	0.6	1.4
	Median	0	1
Not interested/hostility" (%N)	Mean	5.8	15.5
Survey Content /Privacy	Mean	17.6	29.9
Time	Mean	22.7	54.0
Gatekeeping	Mean	3.9	10.7
Other/prior wave	Mean	10.7	26.3

Table 6. R-indicator and response rate from 2 response propensity models

Sample size	No. contact attempts	Base-weighted Response rate	2.5 th percentile	97.5 th percentile	Model DMG			Model DSDMG		
					R-indicator	2.5 th percentile	97.5 th percentile	R-indicator	2.5 th percentile	97.5 th percentile
1,860	1	10.1%	10.1%	10.2%	0.986	0.974	0.994	0.931	0.906	0.954
7,005	3	36.1%	35.8%	36.3%	0.968	0.957	0.978	0.817	0.800	0.832
10,949	5	53.3%	52.9%	53.7%	0.945	0.933	0.957	0.743	0.730	0.757
13,631	7	63.1%	62.6%	63.6%	0.932	0.918	0.943	0.699	0.688	0.712
14,519	8	66.1%	65.6%	66.6%	0.927	0.916	0.938	0.689	0.677	0.701
16,733	12	72.4%	71.9%	73.1%	0.923	0.911	0.935	0.661	0.648	0.671
18,031 (complete Sample)	59	75.2%	74.6%	75.9%	0.923	0.912	0.935	0.642	0.632	0.653

Table 7. Characteristics of reporting behavior of respondents in Wave 1

		No. contact attempts	
		1-7	8+
Sample size (N)	N	11,370	2,197
Introduction to survey letter received (%N)		87.7	91.4
Time to complete survey (minutes)			
Sections 1-20	Mean	40.8	41.3
	Median	36.6	36.1
All sections	*Mean	58.3	66.2
	Median	53.4	59.4
Mode of interviews (CAPI, %N)			
Missing		0.4	0.8
Only visit		85.1	63.6
Only phone		13.4	33.0
Mixed		1.1	2.6
Use of recall aids (%N)			
Missing (infobook and records)		0.4	0.8
1: <50% use/no use of both		47.1	56.2
2: <50% use/no records, >50% use infobk		27.0	22.8
5: >50% use records, <50% use/no infobk		9.3	8.8
6: >50% use of both		16.2	11.4
Unedited expenditure reports			
“Don’t know/refused” values (% reports)	*mean	4.8	6.8
Expenditure on necessities (\$)	*mean	\$1,257	\$1,360
	median	959	1,089
Total expenditure (\$)	mean	\$4,673	\$4,575
	P25	1,450	1,530
	Median	2,811	2,891
	P75	5,217	5,358
	P90	9,502	9,716
<i>Endorsement of filter questions for selected expn categories (%N)</i>			

Table 7. Characteristics of reporting behavior of respondents in Wave 1

	No. contact attempts	
	1-7	8+
4A Telephone services (with cell, internet)	90.7	89.7
4C Utilities	92.7	93.4
6A Major appliances	4.2	4.1
6B Minor appliances	33.5	29.5
7 Maintenance & repairs	25.2	23.5
9A Apparel	55.1	54.6
9B Apparel services	3.9	2.9
I3B Non-health insurance	83.0	80.3
14 Health insurance	70.4	67.6
15B Medical services: payments	48.5	39.3
15C Medical services: reimbursements	1.7	1.3
19 Miscellaneous	53.0	46.4
<i>Combined item reporting (screener endorsed, % endorsers)</i>		
4C Utilities	8.2	7.1
6A Major appliances	17.4	21.1
6B Minor appliances	2.0	2.6
7 Maintenance & repairs	4.3	6.0
9A Apparel	31.7	33.9
9B Apparel services	52.2	62.5
I3B Non-health insurance	6.4	7.3
I5B Medical services: payments	4.1	4.6
I5C Medical services: reimbursements	1.5	0.0
I9 Miscellaneous	1.1	1.7

Table 8. Logistic regression coefficients and significance in final model with specific doorstep concern themes predicting reported (unedited) total expenditures above the median

Predictor	Description	Est. coeff	SE	Wald Chi-sq	P value	Odds Ratio	95LCI	95UCI
Intercept	Intercept	-0.895	0.046	381.551	<.0001			
iatmpgt7	Indicator, >7 contact attempts	0.074	0.060	1.495	0.2215	1.08	0.96	1.21
icntpvgt3	Indicator, >3 visit attempts	0.071	0.049	2.107	0.1466	1.07	0.98	1.18
<i>Specific doorstep themes</i>								
Inh	Not interested / hostility	-0.412	0.080	26.861	<.0001	0.66	0.57	0.77
Icp	Survey Content /Privacy	-0.266	0.053	25.450	<.0001	0.77	0.69	0.85
Itime	Time	0.389	0.046	69.979	<.0001	1.48	1.35	1.62
Igate	Gatekeeping	-0.027	0.090	0.094	0.7596	0.97	0.82	1.16
Iothpw	Other / prior wave	-0.222	0.060	13.882	0.0002	0.80	0.71	0.90
ihmowner	Homeowner	-1.143	0.044	677.225	<.0001	2.23	2.05	2.41
Isingle	CU with only 1 member	0.800	0.042	369.178	<.0001	0.32	0.29	0.35
irecuse	Infobook or Records use >=50% of the time	0.311	0.039	62.186	<.0001	1.37	1.26	1.47
ilenexp	Survey length (sect1-20) >= median	0.996	0.040	616.739	<.0001	2.71	2.50	2.93

N=13,501 (nmiss=66); No. of events=6,769; No. of unique covariate patterns: 1,079

-2 Log Likelihood: 15,867.9; AIC: 15,891.9; R-sq: 0.19; Max rescaled R-sq: 0.25

Predictor	Description	Est. coeff	SE	Wald Chi-sq	P value	Odds Ratio	95LCI	95UCI
Hosmer & Lemeshow Chisq: 11.93, df=8, p=0.15								
Classification: %Concordant=74.8, % Discordant=23.7, ROC area=0.76								

Table 9. OLS regression coefficients and significance in final model with specific doorstep concern themes predicting the number of filter questions endorsed

Predictor	Description	Est. coeff	SE	P value (t)	Standardized coeff	VIF
Intercept	Intercept	4.92	0.04	<.0001	0	0
iatmpgt7	Indicator, >7 contact attempts	-0.14	0.05	0.0098	-0.02	1.34
icntpvgt3	Indicator, >3 visit attempts	-0.12	0.04	0.0067	-0.02	1.28
	Indicator, DS theme:					
Inh	Not interested / hostility	-0.39	0.07	<.0001	-0.04	1.15
Icp	Survey Content /Privacy	-0.19	0.05	<.0001	-0.03	1.18
Itime	Time	0.00	0.04	0.9504	0.00	1.16
Igate	Gatekeeping	-0.10	0.08	0.2232	-0.01	1.06
iothpw	Other / prior wave	-0.01	0.05	0.836	0.00	1.10
isingle	CU with only 1 member	-0.77	0.04	<.0001	-0.15	1.05
ihmowner	Homeowner	0.89	0.04	<.0001	0.19	1.13
irecuse	Infobook or Records use >=50% of the time	0.51	0.04	<.0001	0.11	1.07
ilenexp	Survey length (sect1-20) >= median	1.30	0.04	<.0001	0.28	1.18

N=13,501 (nmiss=66); R-sq and Adj R-sq=0.23; Root MSE = 2.02; Dept Mean=6.05, CV=33.33

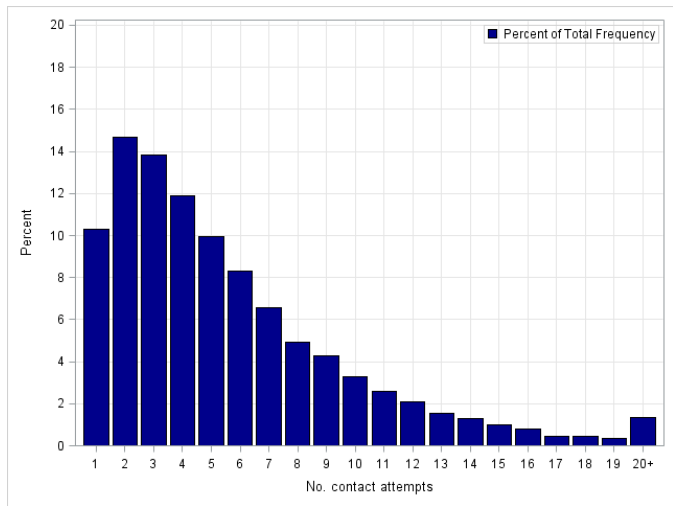


Figure 1a. Overall distribution of contact attempts

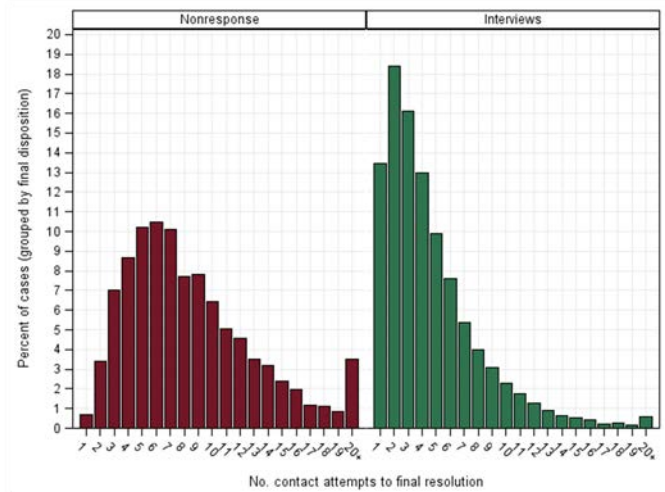


Figure 1b. Distribution of contact attempts by final disposition

