

Optimizing Call Scheduling in an RDD Survey

Whitney Murphy and Colm O'Muircheartaigh, NORC

Summary

The Reach 2010 Risk Factor Survey requires a minimum of seven call attempts to be made over a two-week period in order to complete a household screener. For the first two rounds of the study, calling rules and a detailed schedule were developed to meet these requirements. As the sample for each community was released in batches and sample sizes for later batches were calculated using results from earlier batches, the data collection period for a particular community could sometimes be several months long. Because REACH 2010 collects yearly cross-sectional samples and is interested in charting progress from year to year, it was in the project's best interest to maintain comparability of time periods and compress the data collection period for any one community as much as possible.

After analyzing call history data from the first two rounds of data collection, we decided that reducing the number of days over which the seven calls were made would allow for a shorter overall data collection period for each community without compromising data quality. In the third round of data collection, the calling rules were modified, thus reducing the entire data collection period by almost a third. In this paper, the results of these modifications are compared to the original rules with regard to cost, particularly the average number of calls made to a number, and outcome measures, such as response rates, non-interview rates, working residential number rates, unknown household status rates, and non-household rates. The findings suggest that the changes to the calling rules had no negative impact on cost or outcome while reducing the data collection period.

Background and Problem

\$\$PARAGRAPH ON BASIC IDEAS\$\$

The data come from the second and third rounds of the *Racial and Ethnic Approaches to Community Health (REACH) 2010* survey. REACH 2010 is a demonstration project sponsored by the Centers for Disease Control and Prevention (CDC) to eliminate health disparities among minorities. For this study, NORC conducts yearly sample surveys in 27 REACH communities, 25 of which are surveyed by telephone. This paper concentrates on four telephone communities for which data were available in both the current and previous rounds of data collection. The samples each had dual-frame (RDD and listed telephone number) designs. In each round of data

collection, there were approximately 200,000 calls made to 35,000 telephone numbers across these four communities.

In the REACH telephone communities, NORC is contractually obligated to make a minimum of seven call attempts to complete a screening interview, and these calls have previously been made over a 14-day period. We actually make up to 14 calls to complete a screener and another 14 calls to complete the individual interviews, with each 7-call block spanning 14 days.

During the first round of data collection, numerous research projects were conducted to determine how best to maximize efficiency at the screener level. Murphy, et al. (2003) describes the techniques used to create rules for treating *promising* and *unpromising* cases¹. Using these rules, which were implemented beginning the second round of data collection, all numbers were called seven times over 14 days to complete a screener. Those cases that were not completed after the first seven calls, but were deemed *promising* by the seventh call, were given up to seven more calls over 14 days to attempt to complete the screener. Those cases that were not completed after the first seven calls and were deemed *unpromising* were finalized after the seventh call.

Once the screener was completed, households were given up to 14 calls over 28 days for the individual interviews to be completed. Altogether, a maximum of 28 calls over 56 days were made to a telephone number. Due to call combinations and the rules governing the delays, this time period was often longer than 56 days.

Because the REACH data are used by the CDC to monitor the success of the outreach programs in each community, timing is an important factor. Although the data are not longitudinal, the cross-sectional samples are collected yearly and used to gauge progress; thus, it is in the project's best interest to collect data in each community at approximately the same time each year in order to maintain comparability across rounds of data collection. Furthermore, there are practical issues regarding the

¹ *Promising* cases were defined as those with at least one of the first seven calls resulting in non-negative contact with a household (e.g., an appointment, a request for a letter, or even a soft refusal). *Unpromising* cases were those with negative or no contact with a household, or an inability to determine household status, during the first seven calls (e.g., a combination of ring-no-answers and busy signals, or a hard refusal).

scheduling of interviewers and use of the telephone system that make more concise data collection periods for each community desirable.

Design and Methodology

In the third round of data collection, the calling rules were changed to reduce the time span over which a case is worked. The *promising* and *unpromising* rules were still implemented at the screener level; however, the time period over which each 7-call block was made was reduced from 14 to 10 days. Thus, although the maximum number of calls remained the same at 28, the total time over which those calls were made was reduced from 56 to 40 days.

The primary purpose for this research is to compare the calling rules in the previous round (round 2) to the calling rules in the current round (round 3) to assess the impact that the changes to the rules may have had on the quality of the data and the efficiency of data collection. In order to quantify the impact on quality, we look at several rates at the screener and individual interview level. These rates include the non-household rate, the unknown household rate, the working residential number rate, the refusal rate, and the screener and individual response rates. The impact on cost will be measured by the average number of calls made to a telephone number, which includes calls at the screener and the individual interview levels. A comparison of cost and outcome measures in round 2 to round 3 will help us to determine whether the change in calling rules was appropriate and whether we should continue using these new rules in future rounds.

A secondary purpose for this research was to explore some other potential changes to the calling rules that may be considered for future rounds of data collection. Using the call history data, it is possible to simulate particular changes to the rules, without actually having made them, and assess their impact on data quality and cost. In this paper, we discuss three such simulated calling rule changes.

An analysis that was conducted during round 2 of data collection showed that a large portion of our individual interviews were completed within the first five post-screener calls to a household. This suggested that a reduction in the number of calls at the individual interview level may be warranted, although a change to number of calls was not actually implemented. The first simulation, which we refer to as simulation A, involves reducing the number of calls to a household from 14 to seven. This would further reduce the data collection period from a maximum of 28 calls over 56 days to a maximum of 21 calls over 30 days. The results of this simulation

are compared to the round 3 results using the cost and outcome measures discussed earlier.

A second simulation, simulation B, is a somewhat modified version of simulation A. Under this scenario, we capped the total number of calls to a number at 21, but did not restrict the number of individual interview calls to seven. Thus, if we completed a screener in 3 calls, we would allow up to 18 additional calls to complete the individual interview(s). If, however, it took 14 calls to complete a screener, then we would only allow up to 7 additional calls to complete the individual interview(s).² Thus, this simulation results in a total of 21 calls over approximately 30 days. The results of this simulation are compared to the round 3 results in terms of cost and outcome.

Finally, simulation C builds on the previous analyses that NORC conducted regarding the *promising* or *unpromising* status of a case. Currently, the *promising* rules are applied at the screener level. Simulation C applied a similar notion of *promising* at the individual level. If the individual interviews were not completed within the first seven individual calls, then the history of the case was analyzed and deemed to be *promising* (i.e., worth following past seven calls to complete the interviews) or *unpromising* (i.e., not worth following further and finalized after the seventh call). Again, the results of this simulation are compared to the round 3 results in terms of cost and outcome.

Analyses and Results

Four main analytic comparisons are discussed in this section. The first is the comparison of the round 2 and round 3 calling rules, using cost and outcome measures, to assess the impact of the actual changes made to the rules. The remaining three compare the round 3 rules to the simulated changes, also in terms of cost and outcome.

As discussed earlier, cost is measured by the average number of calls to a telephone number. Specifically, the number of calls will include calls made to complete the screener as well as calls made to complete the individual interview(s). We will be looking at average calls to (1) all numbers, which

² Because of the restrictions in number of calls in our current calling rules, there is still a cap of 14 calls to complete a screener in simulation B. In fact, very few households had more than 14 screener calls. Although there was also a cap of 14 calls at the individual interview level, many more cases had more than 14 individual calls, so it was possible to analyze the effect of going beyond that threshold if the screener was completed in fewer than 7 calls.

include non-households and other cases for which we do not complete a screener as well as households for which a screener and individual interview(s) were completed, and (2) eligible households, which only include cases for which a screener was completed and at least one household member was eligible to complete an individual interview. The overall average will be lower for the first group because of the screening process which weeds out many cases at the screener level, thus making additional individual interview calls to those numbers unnecessary.

Outcome will be measured using several rates, which are described in the previous section. These rates are calculated using standard AAPOR definitions. In all of the following analyses, the *e* used in the screener response rate formula is the working residential number rate, and the *e* used in the individual response rate is the eligibility rate at the interview stage.

Analysis 1: Comparison of Round 2 to Round 3.

For this analysis, we compared cost and outcome measures for round 2 (up to 28 calls over 56 days) to round 3 (up to 28 calls over 40 days). This comparison is used to determine what, if any, impact the change in the number of days over which calls were made had on the quality and efficiency of data collection. Table 1 displays the cost comparison for the two rounds.

Table 1: Mean Number of Calls to a Number, Round 2 versus Round 3

Round	N	Mean	Std. Dev.	Min	Max
Screener Calls (All Cases)					
2	35,248	4.42	3.48	1	18
3	36,923	4.42	3.41	1	19
Screener Calls (Eligible Cases Only)					
2	4,295	3.49	2.72	1	15
3	3,485	3.44	2.74	1	18
Individual Interview Calls					
2	4,295	7.43	5.57	1	23
3	3,485	7.25	5.70	1	41
Total Calls (All Cases)					
2	35,248	5.33	4.52	1	32
3	36,923	5.10	4.25	1	45
Total Calls (Eligible Cases Only)					
2	4,295	10.91	6.43	1	32
3	3,485	10.69	6.52	1	45

The number of calls made in each round at the screener and individual interview stages was very similar. The average number of screener calls to all

cases is identical across the two rounds. There is a small reduction in the average screener calls made to eligible cases, as well as in the average individual and total calls, in round 3, which would suggest that the change in rules actually helped to improve our costs slightly.

Table 2 displays the outcome comparisons for the two rounds. Most of the rates are very similar. The largest difference is in the unknown-household rate, which is quite a bit higher in round 3. This is consistent with the trend toward an increase in unresolved calls (e.g., ring-no-answers and busy signals) that many telephone survey researchers have noticed over the past several years. The refusal rate is also slightly higher in round 3 than in round 2. However, while the screener response rate is one percentage point lower in round 3 (46.12% versus 47.31%), the individual interview response rate is two percentage points higher (61.96% versus 60.04%), which indicates that overall, the change in calling rules may not be harming the quality of the data we collect, and in fact may be improving it slightly. Combining these results with the cost results suggests that the change in calling rules that were made in round 3 were successful and should be continued in future rounds.

Table 2: Rates for Rounds 2 and 3

Rate	Round 2	Round 3
Non-HH	33.91%	32.07%
Unknown-HH	6.70%	13.58%
WRN	63.66%	61.63%
Refusal	27.73%	30.82%
Screener Response	47.31%	46.12%
Interview Response	60.04%	61.96%

Finally, we conducted the cost and outcome analyses individually for each of the four communities being studied here in order to determine whether the trends that we saw overall held across the different areas. While there were some differences among communities (for example, some communities had higher response rates than others), the trends between rounds that we identified overall were also found in at the community level.

Analysis 2: Comparison of Round 3 to Simulation A.

Next, the rates were computed for simulation A, in which individual interview calls were capped at a maximum of 7, thus reducing the total calls to a number from a maximum of 28 over 56 days to a maximum of 21 over 30 days. A comparison of this strategy to the current rules is made to determine

whether this would be an appropriate rule to implement in future rounds of the project.

Table 3 displays the costs for round 3 and simulation A. Because no changes were made at the screener level, the average calls at that stage would be identical under both scenarios, so they are not displayed. As would be expected, the remaining costs are lower under simulation A, given that the individual calls are capped at seven (versus 14 for round 3). The average individual interview calls, total calls to all cases, and total calls to eligible cases are all quite a bit lower under simulation A.

Table 3: Mean Number of Calls to a Number for Round 3 versus Simulation A

Round	N	Mean	Std. Dev.	Min	Max
Individual Interview Calls					
3	3,485	7.25	5.70	1	41
A	3,485	4.69	2.65	1	7
Total Calls (All Cases)					
3	36,923	5.10	4.25	1	45
A	36,923	4.86	3.66	1	25
Total Calls (Eligible Cases Only)					
3	3,485	10.69	6.52	1	45
A	3,485	8.13	3.96	1	25

Given the lower overall costs alone, simulation A looks to be a good alternative to the current rules for round 3. However, the outcomes under these rules also need to be examined. Again, because there are no changes at the screener level, the screener rates are identical for the two groups and will not be presented. Table 4 compares the individual interview response rates for round 3 and simulation A. The interview response rate for simulation A, at close to 53 percent, is almost 10 percentage points lower than for round 3. The cost improvements do not appear to outweigh this dramatic reduction in response rate; thus, we would not recommend the use of this strategy in future rounds of data collection.

Table 4: Interview Response Rates for Round 3 versus Simulation A

Rate	Round 3	Simulation A
Interview Response	61.96%	52.70%

Analysis 3: Comparison of Round 3 to Simulation B.

In this analysis, we examined the effects of capping the total calls made to a number at 21. Under these rules, the individual interview calls were not capped at seven, but instead were capped at the

total calls (21) minus the number of calls required to complete a screener. For example, if a screener was completed in 3 calls, the household would be allowed up to 18 calls to complete the individual interview(s). However, if a screener was completed in 14 calls, the household would only receive 7 calls to complete the individual interview(s). The results of this analysis are presented in Tables 5 and 6 below.

The cost comparison in Table 5 shows that the average number of calls is slightly lower under simulation B, which makes sense given that the total calls are capped at 21 rather than 28 calls. However, this reduction in cost is not as noticeable as with simulation A. This is likely because there were many cases that had more than 7 individual interview calls, and thus would have been affected by the change in rules under simulation A, but there were not as many cases with more than 21 combined screener and interview calls, suggesting that the changes under simulation B would not have as big an impact on cost.

Table 5: Mean Number of Calls to a Number for Round 3 versus Simulation B

Round	N	Mean	Std. Dev.	Min	Max
Individual Interview Calls					
3	3,485	7.25	5.70	1	41
B	3,485	7.02	5.24	1	20
Total Calls (All Cases)					
3	36,923	5.10	4.25	1	45
B	36,923	5.08	4.14	1	21
Total Calls (Eligible Cases Only)					
3	3,485	10.69	6.52	1	45
B	3,485	10.46	5.94	1	21

Table 6 displays the outcome comparisons for round 3 and simulation B. Again, the differences are small. The interview response rate for simulation B is just slightly lower than for round 3.

Table 6: Interview Response Rates for Round 3 versus Simulation B

Rate	Round 3	Simulation B
Interview Response	61.96%	61.20%

Again, we conducted the cost and outcome analyses individually for each of the four communities. The results for the individual communities were almost identical between round 3 and simulation B, thus supporting the overall results.

The fact that there are very small differences in results between round 3 and simulation B is very encouraging. Although we are not reducing our costs or altering the outcome by much, we are reducing the total data collection period for a community from 28 calls over 40 days to 21 calls over approximately 30 days. This is a substantial reduction in time, which is highly desirable from the point of view of the project work, and it comes with no costs in terms of additional work or sacrifice in data quality. Because of these encouraging results, we plan to discuss implementation of these rules in future rounds of the study.

Analysis 4: Comparison of Round 3 to Simulation C.

Finally, through simulation C, we explored the effects of identifying whether households were *promising* or *unpromising* at the individual interview level after the first 7 calls, and then pursuing only the *promising* cases past 7 individual calls. Under this simulation, all screened, eligible households were called a minimum of 7 times to complete the individual interview(s). Then, based on the post-screener call history, the household was deemed either promising or unpromising. If the household was considered to be promising, an additional 7 calls were made to complete the individual interviews. If it was considered to be unpromising, the individual interviews were finalized after the 7th call.

Of the 3,485 eligible households, 1,571 had more than 7 household calls. Of those, only one was found to have an unknown household status, and eight were considered to have unpromising individual contact within the first 7 calls. Thus, the modified rules would only have applied to a maximum of 9 households, or less than 1 percent overall. Given the relatively minute impact any change in rules for those cases would have on the overall data collection cost and efficiency, the simulation was not pursued further.

Conclusions

Several changes in calling rules were investigated in this paper. The change that was made in the third round of data collection, reducing the number of days over which data were collected from a maximum of 56 to 40, appeared to have no negative effect on cost or data quality, and actually seemed to improve both slightly. An investigation into other proposed techniques for further reducing the data collection period was also fruitful. Although we found that a reduction in individual interview calls from 14 to seven would be unacceptably detrimental to the quality of our data, a reduction in total calls from 28 to 21, which has the same benefit in terms of total time in the field for a community, actually reduces costs slightly without negatively impacting data quality. Finally, exploration into a *promising/unpromising* rule at the individual interview stage, similar to the one we apply at the screener stage, proved to affect too few cases to be a worthwhile pursuit.

Suggestions for Future Research

We intend to reproduce these analyses when data are available for all communities, to make sure that the results hold across the other 21 areas. We also plan to investigate the impact these changes have on the demographic make-up of the respondents and the actual individual interview responses.

References

Murphy, Whitney, C. O’Muircheartaigh, C. Emmons, S. Pedlow, R. Harter. 2003. Optimizing Call Strategies in RDD: Differential Nonresponse Bias and Costs in REACH 2010. *Proceedings of the Joint Statistical Meetings, San Francisco, CA, August 2003.*