THE EFFECT OF VARYING THE RESPONDENT SELECTION SCRIPT ON RESPONDENT SELF-SELECTION IN RDD TELEPHONE SURVEYS

Kristina S. Romuald and Lois M. Haggard, University of Utah Survey Research Center
Kristina S. Romuald, Survey Research Center, 2120 Annex Building, University of Utah, Salt Lake City, UT 84112

Key Words: Sampling error, respondent selection, within-unit coverage, last birthday

Abstract. This study examined the accuracy of the last birthday method for randomly selecting respondents within households in a random digit dial (RDD) telephone survey. This study found that informants self-select to participate in surveys at a higher percent (58%) than expected (49%) based on the number of adults in the household and the number of informants age 17 or younger answering the telephone. An embedded experiment was used to examine how a memory cue and/or information provided to informants might affect the probability of informant self-selection. Logistic regression analysis yielded no significant main effects or interactions, suggesting that providing a memory cue or additional information may not alter the probability that the informant will self-select to participate in the survey.

Various methods have been used to randomly select respondents within households for random digit dial (RDD) telephone surveys. According to Maklan and Waksberg (1988) coverage losses in RDD surveys, other than those caused by exclusion of non-telephone households, are primarily due to missed persons in interviewed households. It is important to assess the adequacy of within-unit coverage for each of the various respondent selection methods and to ascertain the best methods for identifying the appropriate target respondent within a household.

One respondent selection technique in telephone surveys has been the Kish method (Kish, 1949) which gives each adult in a household a non-zero and known chance of selection. This method requires a full listing of all adults in the household by age and gender at the beginning of the interview. Based on this information, interviewers use one of six random number tables to identify the appropriate respondent to interview in the household. The advantage of this selection method is that each adult in the household has a non-zero and known chance of selection. According to Lavrakas, Bauman & Merkle (1993), a disadvantage of the Kish method is that it requires interviewers to ask personal household composition questions early in the interview, a process that is both time-consuming and invasive.

In an attempt to simplify the selection process and make the questions less personal, Troldahl and Carter (1964) proposed a method that asks the following questions: 1) How many people 18 years or older live in your household, counting yourself; and 2) How many of them are men? The interviewer then selects the respondent using one of four matrices. Even though this method is less invasive, research has shown that it leads to selection bias in households where there are three or more adults of the same gender, and middle aged people in the household often have no chance of being selected because the "oldest" or "youngest" male or female is chosen. (Lavrakas et al. 1993).

Many researchers now use the last-birthday selection procedure. For this method, the interviewer reads a script such as, "For this survey, I'd like to speak with the person in your household, 18 years of age or older, who had the last birthday" (Lavrakas et al., 1993). This selection procedure has several advantages over other methods: 1) every household member has an equal chance of being selected; 2) someone in the household has to have the next birthday, therefore no working numbers are wasted; 3) members of certain age groups should not be selected more often than others; and 4) potentially sensitive questions are not asked early in the interview (Salmon & Nichols 1983).

A disadvantage to using the last-birthday method is that researchers usually do not ask additional questions about household members' birthdays to verify that the appropriate respondent (the person with the most recent birthday) has been chosen. Also, if the person answering the telephone (the informant) is not the household member with the most recent birthday, s/he may still decide to self-select into the survey sample. In such a case, respondent selection within the household unit is not random. The sample is even less representative if the demographic characteristics of people who normally answer the telephone are significantly different from the demographic characteristics of all household members.

O'Rourke and Blair (1983) conducted a telephone survey in which interviewers asked

respondents to give the month and year of birth for all adults currently living in the household. This question was asked at the end of the survey, and was designed to verify the information given by the informant during the respondent selection process. Results of their study showed that 64.8% of the birthday-method informants were selected as respondents. Data on the household members' birth dates showed that the birthday-selection information given by the informant could be validated in 90.8% of the cases. In other words, it was estimated that nine percent of informants incorrectly identified the person in the household with the most recent birthday. In half of these cases the informants selected themselves incorrectly and in half of the cases informants selected someone else in the household incorrectly. No significant differences in demographic information were found between informants who made correct and incorrect selections within the household.

Although these results are encouraging for the most-recent birthday selection method, more recent research indicates that the percent of respondents being selected incorrectly may be higher. Lavrakas et al. (1993) conducted a similar study in which respondents were asked to name the month in which they were born as part of the demographic section of the survey. Respondents were also asked to identify the months in which all adults 18 years of age or older in their households were born. Results showed that in 25.4% of the cases it was not possible to ascertain whether the correct adult was interviewed because the exact birth date of all household members was not collected, and in 18.5% of the cases the person with the most recent birthday was not interviewed. In summarizing their results, Lavrakas et al. concluded, "This test of the accuracy of the last-birthday method in leading to the selection of the correct adult within the household suggests that in as many as one in four cases, it appears not to work."

Based on 1990 census data of the number of adults in Utah households, the expected proportion of age-eligible informants self-selecting to participate in a survey in Utah is 59% (see Table 1).

Examination of self-selecting patterns in RDD telephone surveys at the University of Utah Survey Research Center (UUSRC) suggested that informants may have been self-selecting to participate in the survey at a higher percent than expected based on the number of adults in the household. Initial observations[1] were based on completed surveys of households with two or more adults. Because informants age 17 or younger should have a zero probability of self-selecting and age of informant was not collected, this initial analysis excluded all households containing children. The expected proportion of self-selecting adults in the sampling pool was calculated at 45%, and the observed rate of completed surveys was found to be 68%.

Table 1. Expected Proportion of Informants Self-Selecting in an RDD Telephone Survey in Utah. (Based on 1990 Census Data for Utah Households)

# of Adults in Hsld.	# of Hslds.	Percent of Total	Expected Percent Self-Selecting	Weighted Percent
1	128,376	24%	100%	24.0%
2	323,740	60%	50%	30.0%
3	57,235	11%	33%	3.6%
4	19,526	4%	25%	1%
5+	6,894	1%	20%	0.2%
Total	535,771	100%		58.8%

Using only completed surveys in the analysis, however, was problematic. When it is the informant who had the most recent birthday, he or she becomes the target respondent and typically goes on the complete the survey. Because repeated call backs are often necessary to complete a survey with noninformant target respondents, this latter group is typically under represented in completed surveys. The result is that the completed surveys yielded a higher rate of self-selection than would have been found for all survey cases.

However, it would also be inappropriate to include all cases in the sampling pool in our calculation of the observed percentage of self-selecting informants. In many cases the informant was under age 18 and should have a zero probability of self-selection. Since data on the age of the informant was not collected, informants under age 18 could not be identified and controlled for in the comparison. To more accurately measure self-selection, data would have to include a measure of the age of the informant.

The first research goal of the current study was to examine the proportion of children (age 17 or under) who answered the telephone so that a more accurate assessment could be made for the expected value of the proportion of informants who should self-select. It was believed that, after controlling for age-ineligible informants in the expected and observed values for informant self-selection, the data would

[1] The original version of this paper, presented at the 1994 annual conference of the American Association for Public Opinion Research, was based on completed surveys only.

show that informants do tend to inappropriately self-select to some degree.

If, in fact, informants do tend to inappropriately self-select to some degree, two possible explanations for this phenomenon seem plausible: 1) that informants purposefully selected themselves out of curiosity, interest, or some other motive, or 2) that informants were not including all adults in the household in their consideration of the last birthday request.

The second research goal was to examine the effect of varying the respondent selection script on informant self-selection. If informants are purposefully self-selecting, the amount of information given to them should influence their propensity to self-select. Informants who are told in greater detail why the last-birthday question is being asked, and that they will be allowed to participate in the survey if they are the household member with the most recent birthday, would be expected to self-select in greater numbers. (In practice, we had tried to remove all such information from the respondent-selection script, but interviewers reported they were more comfortable with a script that delivered information regarding why they were asking the last-birthday question.)

If, on the other hand, informants are inappropriately self-selecting simply because they neglect to consider all adults in the household in their consideration of the last birthday request, one would expect that a memory cue would enhance their accuracy. It was expected that the preliminary question would act as a memory cue, and that informants who received this preliminary question would exhibit a lower self-selection rate.

METHODOLOGY

A statewide RDD telephone survey of 500 Utahns provided the basis for data reported in this study. Two items were included just before the respondent selection script to measure the age and education level of the informant. The informants' age information was used to adjust the expected percentage of self-selectors.

In addition, a respondent selection experiment was embedded in the survey instrument. The purpose of the experiment was to examine the effect of varying amounts of information provided in the respondent selection questions and a memory cue on an informant's decision to self-select into the survey.

Dependent Variable

The dependent variable was whether the informant selected him/herself, or whether the informant selected another household member to respond to the survey.

Experimental Design

Informants were randomly assigned to one of six conditions in a two (memory cue) by three (amount of information) experimental design.

Memory Cue. For half of the cases, the interviewer asked a preliminary question to focus informants on all adult household members. The other half of the informants were not asked the question. This preliminary question read:

> "To start with, how many of the people living at this phone number are age 18 or above?"

Amount of Information. The other independent variable in the design was the amount and type of information provided to the informant in the respondent selection process. The information provided ranged from simply asking for the person with the most recent birthday to explaining why the last birthday question was being asked, and including an additional sentence to make it even more explicit that the person who had the most recent birthday would be interviewed. This experimental design resulted in six versions of the selection questions, which appear in Table 2.

Procedure

All versions of the respondent selection questions were preceded by the following lead-in.

"Hello, this is [fill interviewer name] from the University of Utah Survey Research Center. We are conducting a survey to find out about quality of life and other issues in Utah right now. Your phone number was randomly generated by computer. Is this [fill phone number]?"

Informants were also asked the following questions before hearing the respondent selection items:

- "Is this a residence in the state of Utah?"
- "What was your age on your last birthday?"
- "What is the highest year or grade of school you have completed?

Table 2. Experimental Design

	NOT Preceded by Preliminary Number of Adults Question	Preceded by Preliminary Number of Adults Question
Level 1 Information Provided	"Of the adults in your household age 18 or older, may I please speak to the person who had the most recent birthday?"	"Of these [fill number] adults, may I please speak to the person who had the most recent birthday?"
Level 2 Information Provided	"I need to select an adult at random in your household to interview. Of the adults in your household age 18 or older, may I please speak to the person who had the most recent birthday?"	"I need to select an adult at random in your household to interview. Of these [fill number] adults, may I please speak to the person who had the most recent birthday?"
Level 3 Information Provided	"I need to select an adult at random in your household to interview. We do this by interviewing the person who had the most recent birthday. Of the adults in your household age 18 or older, may I please speak to the person who had the most recent birthday?	"I need to select an adult at random in your household to interview. We do this by interviewing the person who had the most recent birthday. Of these [fill number] adults, may I please speak to the person who had the most recent birthday?"

Informants were then asked one of the six respondent selection questions. Once a respondent was selected, the interviewer read the UUSRC statement of confidentiality and voluntary participation, followed by the survey questions.

The UUSRC utilized a CATI (computer assisted telephone interviewing) system to conduct the interviews. Cases were randomly assigned to interviewers, therefore all interviewers administered all six forms of the respondent selection questions. Interviewer compliance was carefully monitored.

RESULTS

Proportion self-selecting after controlling for age of the informant.

A total of 1022 completed and non-completed cases were identified as eligible Utah households for this survey. A total of 145 of these cases were not included in the analysis because they were hang-ups, household refusals or unresolved callbacks (100), the informant did not answer the age question (32), or the informant did not answer the most recent birthday question (13). The final number of cases included in the analysis was 877.

Before adjusting for informant age the expected value of the proportion of self-selecting informants was 56%. After adjusting that figure for the proportion of age-ineligible informants the expected value was 49% (see Table 3) However, a total of 58% of all informants self-selected to do the survey. A Chi-square analysis revealed that the observed percentage of self-selectors (58%) was significantly higher than the expected percentage of self-selectors (49%) for this sample, ($\underline{X}^2$ (1) = 29.94, $\underline{p} \leq .001$). This suggests that approximately 9% of informants inappropriately self-selected to do the survey.

Table 3. Expected and Observed Percentages of Self-Selectors Adjusted for the Percentage of Minors Answering the Telephone.

	Adults only	Minors only	Adjusted % of self-selectors
Expected % of self-selectors	56%	0%	49%
Observed % of self-selectors	65%	12%	58%

Effect of the memory cue and amount of information on informant self-selection.

The percentage of informants who self-selected to do the survey was analyzed across the six experimental conditions. The percentage of self-selectors for each experimental condition and the number of informants in each treatment condition are reported in Table 4.

Overall, 58% of the informants self-selected to participate in the survey, and 42% selected someone else from the household.

Logistic regression yielded no significant main or interaction effects. Apparently the manipulations had no effect on the probability of informant self-selection.

Table 4. Percentage of Self-Selectors for Each Experimental Condition

	NOT Preceded by Preliminary Number of Adults Question	Preceded by Preliminary Number of Adults Question	Row Totals
Level 1 Information Provided	55% (n=145)	57% (n=136)	56% (n=281)
Level 2 Information Provided	57% (n=154)	54% (n=157)	56% (n=311)
Level 3 Information Provided	67% (n=145)	59% (n=140)	63% (n=285)
Column Totals	60% (n=444)	57% (n=433)	58% (n=877)

Chi-square analyses indicated there was a significant age difference between self-selectors and other-selectors, older informants being more likely to self-select, but this difference was not maintained after controlling for the number of adults in the household. Older informants were more likely to represent single adult households, and were therefore more likely to self-select. There were there were no significant differences between self-selectors and other-selectors across the demographic variables of education level and gender.

DISCUSSION

Results of this study support the theory that informants self-select at a higher rate than expected by chance. The proportion of children (age 17 or younger) who answered the telephone was collected so that a more accurate analysis of the percentage of self-selecting informants could be undertaken. Results showed that informants self-selected at a rate of 58% overall, compared to the expected rate of 49%.

The respondent selection script was varied to explore two possible explanations for informants self-selecting at higher percentages than expected by chance. One explanation was that informants were not including all adults in the household in their consideration of the last birthday request. In an attempt to get informants thinking about all adult household members before the respondent selection question was asked, half of the informants received a memory cue (the preliminary question that asked for the number of adults in the household) and the other half did not. Overall, informants who received the memory cue self-selected at a lower percentage (57%) than informants who did not receive the memory cue (60%), but the difference was not significant. It is possible that the memory cue was effective in reminding informants about all adults in the household, but not effective in getting them to think about the birthdays of those adults. Suggestions for future research include designing a more effective memory cue which also asks informants to think about all household members' birthdays. It would be important to design the memory cue to be effective yet brief, since the purpose of using the last birthday respondent selection method in the first place is that it is less invasive than other methods.

Another possible explanation for informants self-selecting at higher percentages than expected by chance was that informants were purposefully selecting themselves out of curiosity, interest, or some other motive. It was expected that as the amount of information regarding the last birthday request increased, self-selection would increase. Results showed that varying the amount of information given to informants as to why the most recent birthday question was being asked did not have a significant effect on self-selection.

Factors other than the manipulations may have influenced survey results. Introductory information provided about the survey may have sounded interesting to informants, causing them to self-select inappropriately. It may or may not be possible to design more effective versions of the respondent selection questions which would decrease inappropriate self-selection by informants without reducing overall response rates. Perhaps informants who choose to self-select based on the information in the lead-in will self-select regardless of their understanding of random selection within households.

Another factor which may have influenced the results in the experiment is that interviewers asked the informant's age before the respondent selection question. This was necessary for informants under age 18 to be identified so the expected percentage of self-selectors could be adjusted. It is possible that asking the informant's age before the respondent selection questions focused respondents on their own birthdays,

causing them to overlook birthdays of other adults in the household.

The current experiment provided various levels of information explaining why the most recent birthday question was being asked, but did not stress the importance of informants correctly reporting the person with the most recent birthday. It is possible that designing respondent selection questions that focus on the importance of correct identification may lead to a lower self-selection rate. In an effort to actually reduce informant self-selection, perhaps a more effective manipulation could be designed so that informants have a clearer understanding of why it is important to identify the person with the most recent birthday.

In conclusion, future research should focus on designing more effective memory cues for informants, which serve to remind them of all adult household members and their birthdays. Researchers may want to continue research on the amount of information provided and on stressing the importance of correct reporting by the informant.

REFERENCES

Kish, Leslie (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, 44 (September), 380-387.

Lavrakas, P. J., Bauman, S. L., & Merkle, D. M. (1993). The last-birthday selection method and within-unit coverage problems. *Proceedings of the American Statistical Association*, St. Charles, IL.

Maklan, D. & Waksberg, J. (1988). Within-household coverage in RDD surveys. In R. M. Groves, et al. (Eds.) *Telephone Survey Methodology*. New York: John Wiley & Sons.

O'Rourke, D. & Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research*, 20, 428-432.

Salmon, C. T., & Nichols, J. S. (1983). The next-birthday method of respondent selection. *Public Opinion Quarterly*, 47, 270-276.

Troldahl, V. C. & Carter, R. E., Jr. (1964). Random selection of respondents within households in phone surveys. *Journal of Marketing Research*, 1, 71-76.