COST-EFFECTIVENESS AND ADVANCE MAILINGS IN A TELEPHONE FOLLOW-UP SURVEY

William D. Kalsbeek, Scott Joens, University of North Carolina
William D. Kalsbeek, Survey Research Unit, UNC-CH, 730 Airport Road, Suite 107, Chapel Hill, NC 27599

Keywords: Nonresponse bias, survey cost, survey design, geriatric research

I. Introduction

In the current era of shrinking research budgets and decreased sources of funding, many researchers are finding it essential to incorporate techniques from cost-effectiveness theory into their study designs. While every conscientious scientist would like to reduce the expense of conducting his or her research, often this can be a difficult task. This is especially true for researchers who use population surveys as part of their work. When subjects fail to participate in a survey, larger samples must be chosen and higher costs result. Low survey response rates can also cause bias due to nonresponse which can threaten the potential validity of survey estimates.

One way researchers have found to solve the problems of high costs and nonresponse bias is to increase the response rate of the survey. As more subjects participate from a potential sample, the inferential value of the sample increases. In order to improve a survey's initial response rate, then, one must first understand why people choose to (or choose not to) participate in a survey.

Groves, Cialdini, & Couper (1992) proposed a theory that attempts to understand an individual's decision to participate in a survey. The authors' approach is significant because it integrates socio-demographic and survey design factors with social psychological theory to explain the interaction between the interviewer and the respondent. While the Groves, Cialdini, and Couper model was designed for face-to-face surveys, in most regards it can also be applied to telephone surveys. However, one important difference does exist. Since the telephone interviewer does not benefit from the visual cues experienced by the face-to-face interviewer, a significantly different dynamic occurs during the period of the initial interaction. Dillman, Gallegos, & Frey (1976) suggest that these first few moments are crucial, since subjects who continue past this point are almost certain to complete the interview. As a consequence, the success of a telephone interview may be more dependent on the use of preparatory techniques that persuade the respondent to participate in the survey before the initial interviewer/respondent interaction takes place.

One preparatory technique that has been used successfully with mail surveys is the use of advance letters. Schlegelmilch and Diamantopoulos (1991) reviewed 35 studies on the impact of advance letters for mail surveys and found an overall increase of 19 % in the response rates for the advance mail groups. Though the literature on advance letters and telephone surveys is much less extensive (Brehm, 1994; Dillman, Gallegos, & Frey, 1976; Traugott, Groves, & Lepkowski, 1987), similar positive benefits have been encountered (from 6% to 14% increases in response rates for prenotification groups). While advance letters have been shown to produce significant gains in response rates for telephone surveys, these gains do come at a cost. Advance letters result in additional labor, printing, and postage expenses. Furthermore, these costs can increase substantially if the letters are personalized or if they contain incentives for participation.

Even if advance letters are used, does the lag time between the respondent's receipt of the advance letter and his or her first contact with the interviewer affect the response rates? It is generally believed that shorter delays will result in higher response rates, but is this necessarily true? Shorter delays generally mean hiring more interviewers so a greater number of calls can be completed quickly, which results in higher overall study costs. Are these costs justified?

An experiment conducted as part of a national telephone survey of senior citizens will begin to address the following two questions: 1) To what extent will personalization of the advance letter and the inclusion of a modest incentive affect the operational costs and statistical effectiveness of the sample? 2) Will longer delays between the respondent's receipt of an advance letter and the interviewer's first voice contact result in reduced response rates?

II. Elderly Selfcare Study

The 1993-1994 Self-Care Assessment of Community-Based Elderly study was a national telephone follow-up survey of 3,485 non-institutionalized elderly people. The baseline survey conducted in 1991 provided a database on self-care behaviors practiced by the elderly (age 67 and older). The information on self-care behaviors is particularly

important for managing individuals who have functional limitations that might otherwise necessitate institutional care. The follow-up survey of all baseline respondents, conducted in 1993 by the Survey Research Unit at the University of North Carolina, assessed several factors including the extent to which patterns of self-care behavior predicted policy relevant outcomes such as death or institutionalization.

The subjects themselves were interviewed whenever possible. Before the interview began, the subjects were screened for any intellectual impairments with a cognitive screening test. If the subject failed the screening test (about 7% did), a proxy respondent was identified for the interview. Proxies were also used for subjects who were institutionalized or deceased. Almost half of completed interviews in the 1993 survey were proxy interviews.

The sample for the baseline survey was a stratified multi-stage sample that was geographically clustered, with 50 primary sampling units representing both urban and rural areas in the United States. Over-sampling of the oldest age groups was implemented in the final sampling stage so the analyses could focus on the most elderly population. The follow-up survey used a computer assisted telephone interview (CATI) system to collect information about the subjects health, functional status, mobility, and self-care behavior changes from 1991 to 1993.

Prior to the start of the follow-up telephone survey, an advance packet was mailed to each participant to remind them about their participation in the baseline survey and to inform them about the planned follow-up telephone interview. Mailings and the start of calling were staggered into 4 waves (corresponding to 4 random subsets of the baseline sample) to insure timely contact after the subject's expected receipt of the letter. The telephone interviewers used a uniform introductory script for the initial contact with the respondent.

III. The Experiment

The experiment consisted of three different advance packets which varied in the level of personalization for a letter, the level of personalization for a flyer, and the inclusion of an incentive. Each advance letter was carefully designed to maximize its persuasive abilities according to the principles of compliance set forth in the Groves, Cialdini, & Couper model. Personalization of the letter consisted of two levels, either personalized (the subject's name was used in the salutation and embedded in the text of the letter) or non-personalized (no name used -- "Dear study participant"). The flyer had the same two levels, personalized (4 pages with a description of the study, a picture and short biographies of the study participants) and nonpersonalized (2 pages, description only with no picture or biographies). The incentive for the experiment was a refrigerator magnet that thanked the subjects for participating in the study.

All three plans for the experiment contained a letter and one type of flyer. The "Personalized With Magnet" (PWM) plan had the personalized letter, the personalized flyer, and the incentive. The "Personalized, No Magnet" (PNM) plan had the personalized letter and the personalized flyer, but no incentive. The "Non-personalized, No Magnet" (NNM) plan had the non-personalized letter, the non-personalized flyer, and no incentive.

While the study was being conducted, the labor, printing, and postage expenses for each plan were carefully recorded. Over 152 hours of labor were needed to conduct the mailout. As displayed in Table 4, fixed costs included the labor costs for writing the letters, designing the flyers and magnet, and the photo costs for the flyer. Variable costs included labor costs for mailing the letters, postage costs, and printing costs. All costs combined, the PWM plan cost $2.30 per packet, the PNM plan cost $2.05 per packet, and the NNM plan cost $1.23 per packet.

Previous research suggested that the PWM plan would be the most effective, so the three plans (PWM, PNM, NNM) were randomly assigned to study participants in a ratio of 50:25:25, respectively. It was also believed that increased delay in the time between receipt of the advance letter and the initial telephone contact with an interviewer would result in decreased response rates.

Calling began for each wave of the study approximately 7 days after the packets were mailed. When the appropriate subject or proxy was identified by name on the telephone, an assessment was made of the respondent's subjects cognitive functioning before the interview took place. Ten words were read by the interviewer and repeated by the respondent. Respondents who failed the test (3 or more mistakes) were not allowed to complete the interview and a proxy was identified. Seven percent of the subjects failed the cognitive test.

IV. Results

Of the 3,485 subjects in the original survey, 313 were dead-ended before the study started and were not randomized into the experiment. The 313 dead-ends were excluded from all analyses. Table 1 shows the percent participation rates by plan. The response rate represents all subjects who provided completed

surveys. Using the response rates, the three plans differed significantly (chi square = 8.0, $p<0.05$). However, differences in the observed rates were opposite from those one might have expected. The most expensive plan (Personalized With Magnet) had the lowest response rate. Another way to look at these data is to examine the agreement rates. "Agreement" rates represent a respondent's agreement to participate in the study, regardless of whether any data was collected. Agreement rates, therefore, include the 95 subjects who agreed to participate but failed the screen test. Agreement rates were thought to be the best for comparing the ability of the three letter types to persuade recipients to participate in the survey. Findings for this measure (see Table 1) revealed a nonsignificant difference among the three plans (chi square = 1.3, $p = 0.21$). These data suggest that the three plans had equivalent persuasive abilities.

During the course of the survey, some respondents completed interviews even thought their advance mail packets had been returned for incorrect addresses. Since these respondents never received their advance packets, the Chi square analyses were repeated with the return-mail respondents excluded. The results were exactly the same as before. Respondent rates among the three plans showed a significant result in favor of the less expensive packets and agreement rates among the three plans were not significantly different.

Agreement rates by demographic sub-groups are displayed in Table 2. Overall, male gender and rural location had the highest agreement rates for the "All Respondents" columns. A logistic regression model was developed with a 0-1 indicator of agreement to participate as the dependent variable. The number of days delay between the respondent's receipt of the packet and the first interviewer contact was estimated from mail delivery timetables provided by the United States Postal Service. The DAYS DELAY variable was dichotomized into less than or equal to 14 days and greater than 14 days. Less than 14 days was considered a reasonably short delay to examine study participation. The variable PLAN was also dichotomized into personalized (PWM, PNM) and non-personalized (NNM). A stepwise forward method was initially run in SAS (1993) to identify the significant main effects and interactions. Since interactions did not play a significant role, the analysis was then repeated in SUDAAN on main effects only, accounting for the complex sampling design (Shah, 1991). GENDER, LIVING STATUS, and PLAN were all significant at the $p < 0.05$ level. Notably, DAYS DELAY was not significant in the logistic model, which suggests that the variable was not important for predicting agreement rates.

When reviewing the results from Table 2, one should note that the demographic information used in the logistic regression model was based on the subject. However, almost half of the interviews were completed by proxies. For this reason a separate comparison was conducted excluding the proxies (i.e. only using the subjects). The "Subjects Only" columns uses only the subject respondents. Overall, agreement rates for the Subjects Only group are noticeably higher than agreement rates for the whole sample (All Respondents), 89.9% versus 77.6%. A logistic regression in SUDAAN showed only EDUCATION and LOCATION as the significant predictor variables. Neither PLAN nor DAYS DELAY were significant.

The final analysis conducted was to estimate the percent relative bias due to nonresponse. The formula used was based on estimators of the proportion (P) of persons who were impaired, based on various activities of daily living (ADL) scores.

Bias attributed to attrition during attempts to complete a follow-up interview could be estimated directly for measures of the rate of impairment, since the ADL scores had been included in the baseline interview.

Using a weighted estimated of P based on the set of all baseline respondents ($\hat{P}_B$) as the standard for gauging bias, and a comparable weighted estimate using data from only follow-up respondents ($\hat{P}_F$) as the estimate of P, reflecting the effects of follow-up nonresponse, the percent relative bias was obtained as 100 times,

$$\text{Rel} - \text{Bias}(\hat{P}_F) = \frac{\hat{P}_F - \hat{P}_B}{\hat{P}_B} \quad .$$

The weights used to produce both $\hat{P}_F$ and $\hat{P}_B$ were the final adjusted weights computed for analysis of the baseline sample.

The bias estimates in Table 3 were based solely on data for the respondent subjects (i.e., proxy respondents were removed). For PLAN, the bias of estimated impairment is neither consistent in magnitude nor direction. These results would suggest that the levels of PLAN are not predictable for estimates of bias due to nonresponse. The bias estimates for the levels of DAYS DELAY had similar unpredictable bias values.

V. Discussion

Regardless of the type of analysis used, the

results from this experiment suggest that using a personalized advance letter will not improve the response rates for a telephone follow-up survey. Chi square tests were nonsignificant for agreement rates among the three plans, although response rates had significant results but in the opposite direction expected (the most expensive advance letters were the least effective). Moreover, logistic regression analysis of the subject-respondents did not have the variable PLAN (personalized versus non-personalized packets) as a significant predictor of agreement to participate. Finally, bias estimates, through varying in magnitude and direction, were inconclusive for PLAN.

While there was no statistical difference among the three plans, great difference in costs did exist, with the most expensive plan being 87% more costly than the least expensive plan. One must conclude, therefore, that personalization of the advance letter and the use of an incentive are not cost-effective since they significantly increased the operational costs of sample selection but did not increase the statistical effectiveness of the sample. Since personalization of the advance letter did not significantly improve response rates for the telephone survey, the least expensive version of the letter would be considered the preferred version. In addition, the amount of delay between receipt of the advance letter and first telephone contact with an interviewer (DAYS DELAY) was equally insignificant in predicting agreement rates or nonresponse bias.

Why did personalization of the advance letters fail to increase the agreement rate in the follow-up study? The answer may depend on when the respondent's decision to participate is made. If the real decision point is at the time that the interviewer calls rather than when the advance letter arrives, as this study suggests, personalization of the letter may be irrelevant. A personal touch during the telephone call may be sufficient to encourage participation without the need to personalize the advance letter as well. For telephone surveys, advance letters may best be used only as a means to inform the subject about the survey and to prepare him or her for the eventual call.

Dillman, Gallegos, & Frey (1976) explain the importance of preparing the respondent for the interview. The authors believe that unexpected telephone calls catch respondents by surprise and provide them with only a few seconds to make up their minds about participation. The telephone call may startle some respondents or create questions about the legitimacy of the research project for others. A prenotification letter might help to legitimize the survey and alleviate the respondent's anxieties. Petty and Cacioppo (1987) provide a further explanation. When a topic is of small personal importance to an individual (as occurs in most surveys), people will be persuaded to participate on the basis of a heuristic review of the topics extrinsic features, which include interpersonal and societal factors such as the authoritative manner, attractiveness, and credibility of the source. The prenotification letter provides credibility of the source which would encourage participation. One should consider, however, that these results apply to a follow-up survey of the elderly, so our ability to generalize these conclusions beyond this study population and to other types of surveys still remains unclear.

References

Brehm, J. (1994). "Stubbing our toes or a foot in the door? Prior contact, incentives and survey response." International Journal of Public Opinion Research, 6(1): 45-63.

Dillman, D. A. (1978). Mail and telephone surveys: Total Design Method. New York: Wiley-Interscience.

Dillman, D. A., Gallegos, J. G. & Frey, J. H. (1976). "Reducing refusal rates for telephone interviews." Public Opinion Quarterly, 66-78.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). "Understanding the decision to participate in a survey." Public Opinion Quarterly, 56: 475-495.

Petty, R. E., & Cacioppo, J. T. (1986). Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer-Verlag.

SAS Institute Inc. , SAS/STAT User's Guide, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc., 1993.

Schlegelmilch, B.B. & Diamantopoulos, A. (1991). "Prenotification and mail survey response rates: A quantitative integration of the literature." Journal of the Market Research Society, 33(3): 243-255.

Shah, B. V. (1991), "Software for Survey Data Analysis (SUDAAN), Version 6.1, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC.

Traugott, M. W., Groves, R. M., & Lepkowski, J. M. (1987). "Using dual frame designs to reduce nonresponse in telephone surveys." Public Opinion Quarterly, 51: 522-539.

A more detailed version of this paper is available upon request from the authors.

Table 1 PERCENT PARTICIPATION RATES BY PLAN

	Personalized with magnet	Personalized no magnet	Non-personalized no magnet
Response Rate *	75.7	80.9	77.4
Agreement Rate	76.4	79.4	78.5

* Chi Square = 8.0, $p < 0.05$

Response Rate includes only completed interviews
Agreement Rate includes completed interviews and subjects who agreed to do an interview but failed the cognitive screening test

Table 2 AGREEMENT RATES BY SUBGROUP

		All Respondents			Subject Respondents		
			Agreement			Agreement	
Subgroup		n	Rate	p	n	Rate	p
OVERALL		3174	77.6		1350	89.9	
AGE	≤75	1367	78.2		765	90.7	
	> 75	1807	77.2		585	88.5	
RACE	White	2932	77.9		1282	89.9	
	Other	239	75.7		68	86.7	
GENDER	Female	1528	75.1	***	728	88.4	
	Male	1646	80.1		622	91.3	
EDUCATION	≤HS	1343	77.1		502	86.8	**
	> HS	1831	78.1		848	91.5	
LOCATION	Rural	817	80.5		365	92.3	*
	Urban	2357	76.7		985	88.8	
LIVING STATUS	Alone	936	75.6	*	389	86.6	
	Others	2238	78.5		961	91.0	
PLAN	Personalized	2397	77.4	*	1082	90.0	
	Non-personalized	777	78.5		268	88.8	
DAYS DELAY	≤14	1749	78.6		604	90.6	
	> 14	1425	78.3		746	90.4	

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Table 3 ESTIMATED PERCENT RELATIVE BIAS DUE TO NONRESPONSE USING ADL MEASURES

		n	IADL	BADL	MADL	HEALTH
OVERALL		1350	4.10	-0.29	1.06	2.24
AGE	≤75	765	18.50	-2.10	2.29	13.65
	> 75	585	16.63	-2.23	40.33	1.26
RACE	White	1282	18.89	-2.74	17.33	10.65
	Other	68	8.08	14.99	-3.20	3.35
GENDER	Female	728	22.37	-1.21	17.72	-8.24
	Male	622	-0.44	-12.39	6.77	27.27
EDUCATION	≤HS	502	29.15	9.74	47.11	-10.23
	> HS	848	13.00	-6.77	0.85	17.89
LOCATION	Rural	365	20.33	-6.16	17.00	15.62
	Urban	985	13.88	7.29	14.33	-4.20
LIVING STATUS	Alone	389	24.77	4.66	35.72	16.93
	Others	961	14.63	-5.19	5.58	8.36
PLAN	Personalized	1082	16.05	-7.92	21.91	0.20
	Non-personalized	268	25.12	19.87	-6.26	-8.64
DAYS DELAY	≤14	604	24.72	-5.39	31.87	-0.75
	> 14	746	11.06	2.38	-6.28	-2.91

Only Subject respondents are included in this analysis

Table 4 ADJUSTED* COSTS (IN DOLLARS) BY PLAN

Costs	Personalized with magnet	Personalized no magnet	Non-Personalized no magnet
Fixed	$1,137	$1,090	$852
Variable	$2,497	$2,148	$1,092
Total	$3,634	$3,238	$1,944
Cost Per Packet	$2.30	$2.05	$1.23

*Costs adjusted to produce equivalent sample sizes