TOWARD THE DESIGN OF AN OPTIMAL TELEPHONE SAMPLE

J. Burke, D. Morganstein and S. Schwartz, Westat, Inc.

1. Introduction

The survey designer is always faced with the task of optimizing the design in one of two ways: by minimizing the expected error given a fixed budget, or by minimizing the budget given a desired sample accuracy. Traditional sample design texts deal with this problem and provide substantial guidance for developing cost models for area probability samples. With the recent increase in the interest in and use of telephone samples, however, the designer is again in search of information on the relation between the expenditure of resources and the expected accuracy of the survey. This paper discusses a proposal for a cost model and describes the relation between the number of callbacks and the final response rate obtainable as a bias to be included in the error structure. The authors present the results of several recent residential telephone surveys. Several graphical presentations are provided in an effort to display the asymptotic effect of additional callback efforts. Before discussing relevant background studies, a few comments are provided on the expenditure of survey resources to improve accuracy.

It is well known that total survey error contains two components: one attributable to sampling variability and a second associated with nonsampling errors. Whether simple random samples (SRS) or more cost effective clustered samples of phone numbers are selected, the sampling error component will decrease with increasing sample size. For SRS samples, the decrease in variance is of order 1/n. For clustered random digit samples, the sampling error decreases with increasing sample size and with increasing number of clusters in a more complicated way. The relation between increasing sample size and decreasing total error, however, is an asymptotic one. Increasing the sample will reduce the sampling error but not the bias of nonsampling error. One major source of nonsampling error is that produced by nonresponse. The relation between nonresponse and additional callbacks is the subject of this paper.

All surveys operate under the shadow of nonresponse, and all survey designers must decide upon the level of effort to be allocated to its reduction. In field operations, interviewers are instructed as to the number of callbacks households are to receive. In mail surveys, the number of repeated attempts at mailouts to nonrespondents is an important cost component. So, too, does the design of telephone surveys require decisions regarding the number of repeated attempts to be made in determining if randomly selected phone numbers are working residential numbers and then to attempt to obtain the cooperation of eligible respondents.

Below, the authors begin with some recently published results on the effectiveness of callbacks in improving cooperation and completing interviews. After this, a proposal for a cost model is given. An error structure is suggested based on total mean square error and including the possible bias of nonresponse. These results are followed by a review of three recent telephone surveys, wherein the relation between additional callbacks and response rate is described.

2. Background

In this section we discuss two topics. First, we present some previous reports on the relation between callbacks and response rates. Second, we describe three recent Westat projects from which response results were taken.

2.1 Previous Results

There have been several references in recent literature on telephone surveys regarding the relationship of callback to response rates. In Surveys by Telephone, Groves and Kahn describe a study done at the Survey Research Center in 1976 comparing the results of a national sample of households with two telephone samples totaling 12,898 numbers.[1] Their experience with this survey in which one respondent per household was interviewed was that 74.5 percent of working household numbers were closed out after five or fewer calls. To process an additional 15 percent required at least nine calls. Two percent of the number required 17 or more calls. Looking only at households that yielded an interview showed that 78 percent were completed in five or fewer calls.

Jane Williams Bergsten of RTI reported still more favorable statistics in a paper presented at the 1979 annual meetings of the ASA.[2] In a 1978 post-election survey of adult citizens in four western and midwestern states, at least 95 percent of all numbers were successfully classified as eligible or noneligible by the fifth phone call. The survey involved interviewing one randomly selected adult per household. Eighty-six percent of the interviews were completed by the fifth phone call. By the seventh call, the completion rate climbed to 94 percent.

Joseph E. Fitti of the National Center for Health Statistics reported some results from recent Telephone Health Interview System (THIS) surveys.[3] These studies surveyed individuals 17 years or older in telephone households in the 48 contiguous states. There were multiple interviews per household. By the fifth phone attempt, 92 percent of the inscope households had been successfully contacted and 77 percent of those eventually completed had been closed out. Sixty-two percent of the eligible respondents identified were interviewed during the first contact.

2.2 Westat Studies

Below we describe three recent residential telephone surveys. Two of the surveys were national in scope and involved contact by telephone on two different occasions: the first contact was used to collect retrospective information and the latter contact to collect prospective information. These two surveys dealt with traffic accident issues and involved screening of virtually all household members. The third study, conducted in seven areas, was health related and involved an enumeration of household members. A sample of females was subsequently drawn for personal interviewing. Below we discuss each study, describing aspects of the study which might affect response rate.

Survey of Low-Damage Accidents

The survey design for the study involved a national random digit telephone sample of about 12,000 households. Within each household, passenger cars owned by household members were enumerated and the principal driver of each car identified. The drivers were then interviewed at two points in time regarding the involvement of the car in low-speed accidents.

In the first interview, drivers were asked about accidents occurring during the previous six months. The second interview conducted two months after the first, sought information about events occurring between the two contacts.

Address information was requested so that reminder logs could be mailed to the households. The log was to be used to record information on low-speed accidents occurring between the two contacts and to provide a reminder to the respondent useful during the second interview. The initial contact yielded a response rate of approximately 84 percent while the followup contact with the same households resulted in a 98 percent response rate. Most of the small loss resulted from respondents' moving.

Motor Vehicle Accident Survey

Like the survey of low-damage accidents, the design for the motor vehicle accident survey consisted of a combination prospective and retrospective national study. The respondents were people who were found to have been involved in motor vehicle accidents which resulted in personal injury or property damage. The retrospective period for this study was four months; the prospective or followup period was also four months.

Approximately 8,000 households were located via a random digit dialing telephone method. Once identified, household members were screened to detemine if they have been involved in eligible accidents. During the screening interview, every driver in the household was identified and questioned. In addition, the head of the household was identified and asked for information about himself/herself and about nondrivers living in the household. Drivers and household heads were asked if any household members were involved in motor vehicle accidents during the study period and if so, the household member was interviewed regarding the accident.

Heads of households were asked to provide addresses so that a reminder log could be sent in which household members could record the occurrence of an accident. The response rate for the initial contact was approximately 82 percent and for the second contact, 97 percent.

Cancer and Steroid Hormone (CASH) Study

The design for the CASH study consisted of screening by telephone roughly 32,000 households over a 24-month period (about 1,333 households per month) to develop a subsample of female case controls. Households were located via a random digit dialing telephone method in seven prespecified locations drawn from a mix of urban and rural areas. During the screening interview, Westat collected age, sex, and address information for each household member between 20-54 years. The screening information was provided by an adult member of the household. Westat then subsampled female cases who were personally interviewed at their homes. The study's response rate, to date, has been 92.0. The data reported in this paper are for the first four months of the telephoning period. The average duration of telephone contact is four minutes.

3. The Optimization of Survey Design

The selection of design parameters is based on an examination of costs and on an understanding of the error structure. Either costs are minimized for a desired level of precision, or total error is minimized for fixed costs. Often, the multiple objectives of the survey are collected together in a somewhat imprecise way and one or more statistics "lumped" together as a focus for the development of the error model. That is to say that, it is generally not possible to minimize the errors of several variables at once, rather to seek an overall, roughly optimal, solution. Fortunately, the solution is rarely so sharply defined as to make this approach unworkable.

To proceed along these lines, a cost model and a specification of the error structure are needed. Below, we suggest a fairly straightforward model for the allocation of costs. After this, a discussion of errors is given. The survey accuracy includes a term for the possible bias of nonrespondents. We suggest that the mean square error, including a bias term, be used in the "optimization" conducted to select design parameters since the nonresponse rate of phone surveys cannot be expected to be trivial.

3.1 Development of a Cost Model

To develop a cost model we will first review the steps of a random digit dial telephone sample involving screening of households. With this process in mind, the elements of cost will be clearer. Further, the Waksberg method of cluster sampling, commonly used in RDD surveys, will be included in the design.

The first step of the process is to identify residential clusters. To accomplish this, a random four digit number is appended to a randomly selected six digit telephone area code and exchange. The phone number is called. If it is a residential number, the initial eight digits are retained and form the seed for additional calls. This is, in effect, a screening process whereby phone clusters without residential numbers are removed from the sample. Thus, if a group of 100 phone numbers, identified by the same first eight digits, have no working residential number, then the group will receive at most one call and then be dropped from the sampling operation.

After the identification of clusters, a predetermined number of calls to residential numbers are made, thus yielding equal size clusters. A thorough discussion of the probabilities of selection and the selection of cluster size are given in Waksberg (1978).[4] At each of the residences located, the initial contact is made and the household screened for eligibility in the survey. (While many surveys may consider all households eligible, the inclusion of this step provides more generality.)

Once the households eligibility has been established, the appropriate respondent(s) is contacted and interviewed. This step may require a substantial number of callbacks, particularly if more than one respondent is required. In two of the surveys we are describing, more than one respondent was interviewed.

Given the above description of the process, we define the following cost parameters:

C_o = the fixed cost of the survey, independent of sampling;

C_{id} = cost to make one contact for the identification of clusters;

K_{id} = the number of additional callbacks to identify clusters;

C_{sc} = cost to make a screening contact;

K_{sc} = the number of additional callbacks to complete residences;

C_{in} = cost to complete one interview;

K_{in} = the number of additional callbacks to complete the interview;

$\bar{q}$ = the average number of main interviews to be conducted per household;

n = the number of residences per phone cluster;

m = the number of phone clusters;

f_1 = the proportion of phone numbers in the given universe expected to be residential (for a national survey, this is approximately .2); and

f_2 = the proportion of households expected to be eligible for the survey.

Given the above definitions, one possible model for overall survey cost is given by:

$$C = C_o + C_{id}(1+K_{id})m/f_1 + C_{sc}(1+K_{sc})mn + C_{ij}(1+K_{in})mn\bar{q}f_2 \qquad (1)$$

For fixed numbers of callbacks, this cost function can be rewritten in the form:

$$C' = m(C_1 + C_2 n) \qquad (2)$$

as used by Waksberg (1978). When the number of callbacks are fixed, the determination of optimal cluster size can be made using the results given in the referenced paper.

In expression (1) the cost of conducting a callback is the same as that of the interview or screening. The expression allows for a different number of callbacks at each step. A second cost model requiring the same number of callbacks, k, at each stage but breaking out a separate cost per callback, C_{cb}, is given by:

$$C = C_o + (C_{id} + kC_{cb})m/f_1 + (C_{sc} + kC_{cb})mn + (C_{in} + kC_{cb})mn\bar{q}f_2 \qquad (3)$$

We believe that expression (3) is better for use in representing random digit dial telephone survey. The requirement of the same number of callbacks at each stage presents no operational difficulties while resulting in a much simpler analytical problem. In this form, there are only three unknown constants, k, m, and n. Below, we propose an error model containing these constants and describing their effect on sampling errors and on bias.

3.2 An Error Structure

The error model of the RDD telephone survey contains at least the error of a simple single stage cluster design and a term summarizing the possible biases of nonresponse, that is:

$$\text{mean square error} = \frac{\sigma^2}{mn}[1 + \rho(n-1)] + \text{Bias}^2 \quad (4)$$

where σ^2 equals the unit variance of a simple random sample; and ρ equals the intraclass correlation among clusters.

In the next section of this paper we provide some indication of the form of the possible bias term. We propose to represent these curves as simple exponential functions of the number of callbacks, with an asymptotic value of zero. Of course, nonresponse cannot be totally eliminated even with a large number of callbacks, but we will assume so for the present. Thus, the bias will be of the form:

$$\text{Bias} = a \cdot \exp(bk) \quad (5)$$

where a and b are some appropriate constants and k the number of callbacks for the particular step in the process: cluster identification, screening or interviewing. For more generality, an additive constant can be included in (5) to represent the nonresponse which remains regardless of additional callbacks.

The bias associated with stopping after k calls can be approximated by the number of households not completed after the kth call which ultimately would have received a particular status had a sufficient number of phone attempts been made. Expression (5) can be linearized by taking the natural logarithm of both sides of the equation. This yields:

$$\text{Ln(Bias)} = \text{Ln}(a) + bk \quad (6)$$

Fitting straight lines to the transformed data yields r^2 values in the range of .89 to .99, indicating a strong linear relationship. Table 1 displays the value of r^2 along with estimates of slope and intercept parameters and their standard errors for the various final status codes within each study. An analysis of covariance within each status indicates that the slopes are not equal across studies.

By combining the cost models and error structures given above, an optimization problem is posed.

4. The Number of Calls Needed to Complete a Case

Below we describe four types of finalized statuses: nonresidential numbers; nonworking numbers; residences yielding a completed interview and residences which ultimately refuse to cooperate. The data presented for the three phone surveys consist of cumulative distributions of the number of calls needed to finalize the case. The number of cases unfinalized at any point in the survey may be viewed as a potential bias of the survey results if interviewing were to stop at that point.

The general form of the curves may be modeled as exponential functions. Once parameters have been estimated, such functions can be used to represent the bias component of the total survey mean square error, as described in Section 3.

The four types of finalized statuses are summarized in Figures 1-4 below. We should mention the difficulty of generalizing survey results and response patterns. The three surveys reported on below are somewhat different in nature having telephone contact as a common denominator. We believe, however, that an examination of such data will be beneficial to telephone survey designers.

4.1 Nonresidential

Figure 1 shows that of the three surveys, the Driver Survey identified its nonresidential numbers with the fewest number of phone calls. By the seventh call, 100 percent of those numbers receiving a final disposition of nonresidential had been identified. The MVA survey, whose nonresidential numbers took the longest to resolve, had finalized over 90 percent of the nonresidential numbers by the seventh call.

4.2 Nonworking Numbers

As seen in Figure 2, a different pattern emerges for nonworking numbers. Seventy to ninety percent of nonworking numbers were identified on the first telephone attempt. This is because a large percentage of such numbers have a recorded message identifying them as nonworking. By the fourth call, over 95 percent of the numbers had been resolved in all three surveys.

4.3 Completes

The data relating to number of calls required to finalize a cooperating household show the greatest consistency among the three surveys. We see in Figure 3 that by the fifth phone call over 90 percent of the completes had been finalized in each of the surveys. The CASH survey met with faster completion primarily because of the brevity of its screening questionnaire as compared to the lengthy interviews used in the other two surveys. By the ninth call, 97 percent or more of the completes were finalized.

4.4 Refusals

In Figure 4 are data pertaining to refusals. In only a small fraction of the cases was a household classified as a refusal on the first phone call. In most instances, at least one other contact was made in an effort to convince the potential respondent to cooperate. All three curves in the figure exhibit a relatively steep slope before leveling off at about the seventh call.

This indicates that each additional callback results in a significant return in finalizing refusals. This is in contrast to some of the earlier figures where a more gradual slope shows that after a certain number of attempts, additional callbacks did not yield many more finalized cases.

5. Summary

In this paper we suggested cost and error models for use in optimizing a telephone sample design, particularly as relates to the number of callbacks to be attempted. We proposed the minimization of total means square error. Further, we presented results from three recent surveys to help quantify the reduction in nonresponse brought about by additions calls, a relation needed in the optimization process. Further work will be aimed at estimating parameters needed in the cost and error models and at finding an optimal number of clusters, cluster size, and number of callbacks required.

Footnotes

[1]Robert M. Groves and Robert L. Kahn, Surveys by Telephone, New York, Academic Press, 1979. pp. 37, 55-59.

[2]Jane William Bergsten, "Some Methodological Results from Four Statewide Telephone Surveys Using Random Digit Dialing", ASA 1979 Proceedings of the Section on Survey Research Methods, pp. 239-243.

[3]Joseph E. Fitti, "Some Results from the Telephone Health Interview System", ASA 1979 Proceedings of the Section on Survey Research Methods, pp. 244-249.

[4]Joseph Waksberg, "Sampling Methods for Random Digit Dialing" Journal of the American Statistical Association, March 1978, Volume 73, Number 361, pg. 40-46.

Table 1. Estimates of regression parameters for equation Ln(Bias) = Ln(a) + bK

		Completes		Final refusal		Nonresidential		Nonworking	
		Estimate	Std. error	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
	b	- .6105	.0093	- .5774	.0438	- .5136	.0179	- .4665	.0534
CASH	Ln(a)	8.2112	.0603	6.1843	.3097	6.4194	.1159	6.3689	.3158
	r^2	.9977	--	.9405	--	.9881	--	.8946	--
	b	- .7614	.0259	- .8458	.0637	-1.1789	.1703	-1.0351	.0622
DRIVER	Ln(a)	9.0042	.1684	7.1805	.2663	7.7951	.6141	8.1762	.2961
	r^2	.9885	--	.9671	--	.9055	--	.9754	--
	b	- .5354	.0064	- .6824	.0396	-1.0498	.0852	-1.2615	.1362
MVA	Ln(a)	8.8130	.0602	7.5166	.2343	7.5600	.3072	8.3364	.4124
	r^2	.9979	--	.9706	--	.9681	--	.9554	--

Figure 1. Percent of nonresidential numbers identified by number of telephone calls

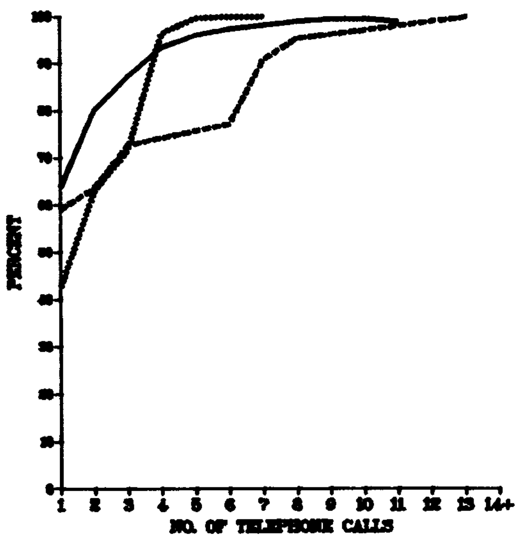

Figure 2. Percent of nonworking numbers identified by number of telephone calls

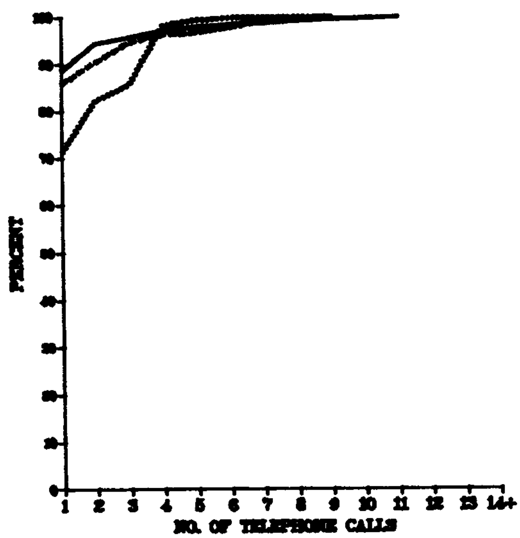

Figure 3. Percent of households completed by number of telephone calls

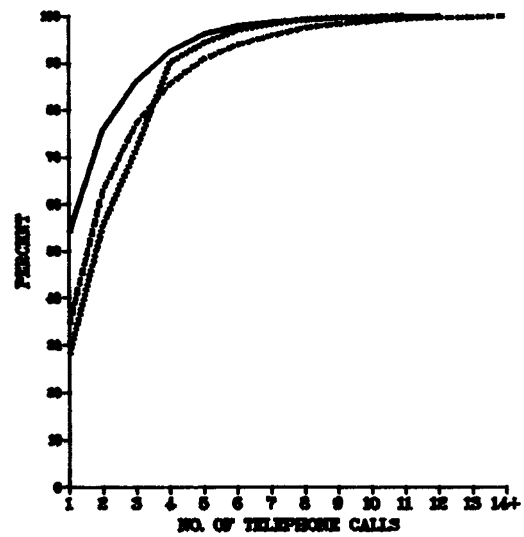

Figure 4. Percent of refusals finalized by number of telephone calls