An Application of a Method of Determining Optimum Number of Call Attempts in a Telephone Survey

K.P. Srinath, Pamela Giambo, Susan Arday
Abt Associates Inc. and Centers for Medicare and Medicaid Services

Introduction

In recent years, nonresponse to data collection attempts in telephone surveys is increasing due to a variety of reasons. To increase the response rate and reduce possible bias due to nonresponse, a common practice is to make several call attempts to collect data from eligible households or persons. These additional attempts increase the survey cost. Also, there cannot be an unlimited number of attempts to collect data as the total survey cost is generally fixed. Therefore, it is important to determine whether the reduction in bias and mean squared error of the estimate resulting from an increase in the number of call attempts is worth the increase in the cost of the survey. In other words, we want to determine the optimum number of call attempts taking both cost and the reduction in bias of the estimate into account. There are several approaches to determining the optimum number of call attempts. For example, one approach is to focus on minimizing the bias due to nonresponse. An alternative approach is to adjust the mean squared error after each attempt for the increased cost of collecting data and compare the adjusted means squares (Srinath et al., 2002). We apply this method of minimizing the adjusted mean squared error to the Immunization Remeasurement Survey conducted in 2002 by the Centers for Medicaid and Medicare Services (CMS) under a contract with Abt Associates Inc. to estimate the influenza and pneumonia vaccination coverage rates among the Medicare beneficiaries.

In section 2, we provide a brief description of the Immunization Remeasurement Survey. In section 3, we describe the method of adjusting the mean squared error after each call attempt for the increased cost of collecting the data to determine the optimum number of call attempts. . In section 4, the results of the application of this method to the survey are discussed.

2. Description of the Immunization Remeasurement Survey

The objective of the survey was to estimate coverage rates of the influenza and pneumonia vaccinations for the Medicare beneficiaries. A random sample of Medicare beneficiaries was selected in each of the 50 states and the District of Columbia and Puerto Rico. The data was collected in two waves by telephone. Wave 1 of the survey was done in 2001 and included 35 states and the District of Columbia and wave 2 of the survey was done in 2002 in the remaining 15 states and Puerto Rico.

The sampling frame for the selection of Medicare beneficiaries was the Enrollment Data Base (EDB) maintained by the CMS. The EDB contains the names of the beneficiaries, addresses and some basic demographic data like age, gender and race/ethnicity. Samples from the EDB can be selected using unique identifiers like Health Insurance Claim (HIC) number. For the first phase sample selection, a 5% systematic sample was selected in each state and the two regions. All persons in this 5% sample who were identified as living in a nursing home, younger than 65 years, not enrolled in either Medicare part A and/or part B were removed from the first phase sample. The list constructed using the first phase sample was the sampling frame for the selection of the second phase sample. The list was stratified by into three age strata, which were 65-74, 75-84 and 85+. The sample was allocated to each stratum roughly in proportion to the number in the first phase sample. Within each stratum, a systematic sample of persons was selected after sorting by race, gender and age. It was required to collect data from 500 beneficiaries in each state. A sample larger than 500 was selected to allow for nonresponse and ineligibles. The sample for data collection was released in replicates.

The questionnaire used for data collection was a CMS-modified version of the 1999 Behavior Risk Factor Surveillance Survey (BRFSS) conducted by the Centers for Disease Control (CDC) which included questions on influenza and pneumococcal pneumonia as well as demographic information needed for analysis.

As a part of the design, a pre-notification letter was s mailed to all selected beneficiaries stating the reasons for the study and saying that they would be called. Since the EDB does not contain telephone numbers and this survey as conducted over the telephone, telephone numbers for the selected sample were generated using several approaches. Data was collected using computer-assisted telephone interviewing system (CATI) from centralized telephone interviewing facilities. As indicated earlier, the objective was to obtain data from a sample of 500 beneficiaries in each state and Distract of Columbia and Puerto Rico.

Calls were made to the selected beneficiaries from whom telephone numbers available. The CATI system kept track of the results of each call by assigning disposition codes. For example, the first call to a selected beneficiary could result in data collection, or refusal, or some other result needing a callback. Maximum number of call attempts was set at 24 for telephone numbers that resulted in non-contact with the selected beneficiary. Therefore, data from the respondents in the survey may have been obtained after the first or second or third etc. call attempts. A nonrespondent after the first call attempt may have been classified as a respondent after the second call attempt. Some examples of disposition codes after several attempts are (1)complete, (2) refused, (3) answering machine, (4) call back, (5)deceased.

3. Optimum Number of Call Attempts

As indicated earlier, in this paper, we examine the bias versus cost trade-off in limiting the number of call attempts. For example, after the first call attempt, we look at the disposition codes assigned to the sample and classify persons as respondents, nonrespondents and ineligibles. We use nonresponse-adjusted sampling weights to get an estimate of the population parameter of interest after each call attempt.

The method described below is taken from Srinath et al. (2001). We are interested in estimating the influenza vaccination coverage rate for the Medicare beneficiary population. The method suggested here is a simpler version of the method of comparison used by Deming (1953).

Let Z denote the influenza vaccination coverage rate in the population of Medicare beneficiaries. Let $t = 1.2, \ldots . A$ denote the number of call attempts. A denotes the maximum number of attempts to collect data before designating a person as a nonrespondent. Let z_t represent the estimate of the population vaccination coverage rate Z based on n_t completes after t attempts. Let $E(z_t) = Z_t$. Therefore, the bias in the estimate z_t is $(Z_t - Z)$. We expect z_t to have a smaller bias due to nonresponse when t is large and a larger bias when t is small. We assume that z_A based on maximum number of call attempts and the largest number of completes n_A to be unbiased.
That is, $E(z_A) = Z_A = Z$. Only the estimate based on the maximum number of attempts is assumed to be unbiased.

Let the variance of z_t be $V(z_t)$. The mean squared error of z_t is given by

$$M(z_t) = V(z_t) + (Z_t - Z)^2 .$$

Let the observed cost per completed interview after t attempts be c_t. The total cost of the survey after t attempts is $c_t n_t$. We have $c_t n_t > c_{(t-1)} n_{(t-1)}$ because of greater number of completes and increased cost per completed interview. Also, we have

$V(z_t) < V(z_{(t-1)})$ and generally expect $(Z_{(t-1)} - Z)^2 > (Z_t - Z)^2$. This decrease in bias and variance is at an increased cost. We have to look at the decrease in mean squared error holding the cost fixed. We adjust the mean squared after each attempt for a fixed cost and compare the adjusted mean squared errors. The adjustment is based on computing two sets of expected number of completes after each attempt, one set assuming fixed total cost and the other set assuming an increase in cost for later attempts.

Let C denote the total budget available for data collection. If the cost per complete after one attempt is c_1, then the expected number of completes after one attempt is $m_1 = \frac{C}{c_1}$ if C is fixed. If the response rate to the first attempt is r_1, then to get an expected sample of m_1 completes, we need to select an initial sample of $m_0 = \frac{m_1}{r_1}$ Medicare beneficiaries. If we start with an initial sample of m_0 Medicare beneficiaries and apply expected response rates at various attempts, we get the expected number of completes at the end of

these attempts. If the response rate at attempt t is r_t, the expected number of completes after t attempts with an initial sample of m_0 beneficiaries is $m_t = m_0 r_t$.
If the cost per complete after t attempts is c_t, then we have $c_t n_t > C$. If we want to keep the total cost fixed, then we can only have $m_t = \frac{C}{c_t}$ completes. Let $f_t = \frac{n_t}{m_t}$. f_t represents the ratio of the expected number of completes with increasing cost to the expected number completes with a fixed cost. We use the ratio f_t to inflate the variance of the estimate based on n_t completes after t attempts. This inflated variance is used to compute the adjusted mean squared error.

The adjustment ratio f_t can be written as $f_t = \frac{n_t}{m_t} = \frac{m_0 r_r}{C / c_t} = \frac{r_t c_t}{r_1 c_1}$ since we have $C = m_1 c_1$ and $m_0 = \frac{m_1}{r_1}$. The ratio for the first attempt f_1 is equal to 1 since we consider one attempt as the minimum number of attempts. If we consider 3 attempts as the minimum number of attempts, then the adjustment ratio will be

$$f_t = \frac{r_t c_t}{r_3 c_3}.$$

We want to identify the number of call attempts $t = 1, 2, ...A$ that minimizes the following adjusted mean squared error of the estimate after t attempts.

$$Mse(z_t) = V(z_t)\frac{r_t c_t}{r_1 c_1} + (Z_t - Z)^2.$$

Since $V(z_t)$ and $(Z_t - Z)^2$ are unknown, we minimize the estimated mean squared error

$$mse(z_t) = v(z_t)\frac{r_t c_t}{r_1 c_1} + (z_t - z)^2.$$

4. Application to the Immunization Remeasurement Survey

We computed the estimates, the bias in the estimates, variance and the adjusted means squared error in the Immunization Remeasurement Survey at 1, 4, 5,6,7,8 12, 15, 20 and 24 attempts. The estimate obtained after 26 attempts were considered as unbiased and bias after a given attempt was estimated as the difference between the estimate after the attempt and the unbiased estimate. We also evaluated the response rates and cost per complete after various attempts mentioned above. Table 1 gives the ratios of the response rates and cost per complete needed to adjust the mean squared error. Table 1 shows the bias in the estimates of vaccination coverage rates and the variance and the adjusted variance and the mean squared error.

Table 1: Ratios of Costs per Complete and Response Rates for Various Attempts

Call- Attempt t	**Ratio of Response Rates** $\frac{r_t}{r_1}$	**Ratio of Cost per Complete** $\frac{c_t}{c_{1\ t}}$	**Adjustment Factor** $\frac{r_t c_t}{r_1 c_1}$
1	1.000	1.000	1.000
4	2.476	1.318	3.263
5	2.709	1.395	3.779
6	2.885	1.459	4.209
7	3.014	1.516	4.569
8	3.107	1.565	4.862
12	3.311	1.702	5.635
15	3.399	1.777	6.040
20	3.471	1.860	6.456
24	3.524	1.986	6.999

From Table 1, we see that the adjustment factors increase because of increased cost and response rates in later attempts.

Table 2 shows the estimated influenza vaccination coverage rate after each of the stated call-attempts, the bias in the estimates, variance and the adjusted mean squared error. From Table 2, we see that the adjusted mean squared is minimum for 6 call attempts. After this there does not seem to be much gain in relation to increased cost.

Table 2: Bias and Mean Squared Error of the Estimates from the Immunization Remeasurement Survey

Call-Attempt	Estimate	Bias in the Estimate Percentage Points	Variance	Adjusted Variance Variance x Adjustment Factor	Adjusted Mean Squared Error
1	73.24%	1.86	0.5853	0.5853	4.045
4	72.50%	1.12	0.2226	0.7263	1.981
5	71.88%	0.50	0.2082	0.7868	1.037
6	71.66%	0.28	0.1956	0.8233	**0.902**
7	71.70%	0.32	0.1854	0.8471	0.952
8	71.67%	0.29	0.1790	0.8703	0.954
12	71.53%	0.15	0.1668	0.9399	0.962
15	71.59%	0.21	0.1620	0.9785	1.023
20	71.50%	0.12	0.1589	1.0258	1.040
24	71.38%	0.00	0.1569	1.0981	1.098

References:

Srinath, K.P., Battaglia, M.P., Cardoni, J., Crawford, C., Snyder, R., Wright, R.A. (2001). Balancing Cost and Mean Squared Error in RDD Telephone Surveys: The National Immunization Survey. Proceeding of the Survey Research Section Methods Section, American Statistical Association.

Deming, W.E. (1953). On a Probability Mechanism to Attain Economic Balance Between the Resultant Nonresponse and the Bias of Nonresponse. Journal of the American Statistical Association, 48.