SAMPLE SIZE VERSUS CALL-BACK EFFORTS: AN EMPIRICAL APPROACH

David J. Fitch, National Center for Health Services Research and Health Care Technology Assessment

In designing the strategy to be used in a sample survey, decisions must be made on the allocation of available funds between those used to collect data from people more easily located and interviewed, e.g., people who are found to be at home on the first call and agree to be interviewed, and funds used to support an effort involving, perhaps, multiple call backs and special efforts to enlist the cooperation of those people who initially decline to be interviewed. The cost per completed interview with people of the first type will obviously be less than interviews of the latter type which are more difficult to obtain. To the extent that the data collection funds are used to interview people of the first type, and not the second, it will be possible, given a fixed budget, to interview more people. The larger the group interviewed the smaller we can expect our sampling variances (standard errors squared) to be. On the other hand one would expect those located and interviewed without call-backs would be at least somewhat different than those who are part of the sample drawn but who can not be located and interviewed on the initial call. To the extent that there is a difference in how those in our sample who are interviewed respond, and the responses we would have gotten from the remainder of our sample had we been able to interview them -- to this extent the means which we calculate from those interviewed will be biased. We call this response bias, or more accurately, nonresponse bias. The Mean Square Error, MSE, is then the sum of the sampling variance, $V(\bar{y})$, and he bias squared:

$$(\bar{y} - \mu)^2$$

It is a measure of the goodness of our obtained mean as an estimate of the population mean.

So to repeat, in designing a survey one is faced with the question as to how to allocate data collection dollars between increasing the number of people interviewed, which would be expected to reduce the sampling variance, and putting additional dollars into call-back efforts which would be expected to reduce the nonresponse bias. The goal is to minimize the MSE, a function of both sampling variance and bias.

There would seem to be at least two quantitative approaches to this problem of resource allocation. One could, based on the large amount of available research and experience, develop a model of expected responses in the population to be sampled. This would be primarily with regard to means and variances of difficulty-of-finding-at-home-and-successfully-interviewing subpopulations plus the costs of locating and interviewing each such group. These models could then be used to make estimates of the optimum number of call-backs. This is the approach taken by Deming in his classic article (1953). The definitive work here is by Rao (1966, 1982, 1983).

The approach in what is reported here is to work with data from a completed survey in which there is information available on the amount of effort which was made to locate and interview sample members so that those interviewed can be divided into groups, the placement of each interviewee depending upon the amount of effort expended to locate and achieve his or her participation. I had hoped that tape files from the Household Interview component of the National Medical Care Expenditures Study (NMCES) would include information on the number of calls made before the successful interview but this is not the case. It is hoped that the files for the new National Medical Expenditure Study (NMES) will include such information. What are available, with some information on effort before a successful interview, are data from the Physicians' Practice Survey (PPS) component of the NMCES. For information on this survey and efforts made to obtain a high rate of participation as well as analyses very relevant to the present study, see Berk (1983).

In this Physicians' Practice Survey, using a 15 minute telephone interview, efforts were made to collect data from a sample of about 6,700 physicians. We can describe the data collection, somewhat oversimplified, as follows. During months 1-2 efforts were made to interview this sample of physicians with 1,762 successes. We will call these type 1 cases. Type 2 cases are the 1,517 additional successes in months 3-4. During months 5-6 further additional efforts were made to interview those physicians still not yet surveyed. This resulted in the final 1,666 cases which we label type 3. As a great deal of effort, i.e. funds, went into call backs and other special attempts to get interviews during the second and third time period, the question arises as to whether these data collection funds, which could be expected to reduce bias, might have been more profitably spent to interview more physicians of the type successfully located and interviewed with less effort and expense during the first time period. This would have increased the total number of cases over that total actually achieved and we would thereby expect a reduced final sampling variance component of the MSE. As there are three different call back types, we have a chance to ask our question twice, i.e., we can ask what the data indicate the effect on our MSE's would have been (1) had we limited our interviewing to type 1 cases and (2) if we had limited ourselves to type 1 and 2 cases.

It should be noted that, although complex sampling procedures were used, this was not taken into account in estimating sampling variances, i.e., computations correct only for the case of simple random sampling were used. Neither were corrections for finite populations employed.

Specific Computations

These notes following, although incomplete, will likely suffice for understanding what follows.

n_1, n_2, n_3 The number of type 1, 2, and 3 cases, i.e., physicians of the type interviewed during the 1st, 2nd, and 3rd time periods.

$V(\bar{y}_1)$ Estimate of sampling variance for some variable y of interest from the PPS based on type 1 cases.

$$V(\bar{y}_1) = \sum_{i=1}^{n_1} \frac{(y_{1_i} - \bar{y}_1)^2}{(n_1 - 1)\, n_1} = \frac{s_1^2}{n_1}$$

$V(\bar{y}_{1+2})$ The sum subscript designates a statistic based on pooled cases, in this case type 1 and type 2 cases.

n_1' The modified number of type 1 cases that would be expected to yield a modified sampling variance designated as $V(\bar{y}_1)'$.

The following sequence of calculations was undertaken for a number of variables from the PPS, each represented by y.

$$MSE_{1+2} = V(\bar{y}_{1+2}) + (\bar{y}_{1+2} - \hat{\mu})^2$$

If

$$MSE_{1+2} > (\bar{y}_1 - \hat{\mu})^2$$

i.e., if it might be theoretically possible to have obtained a sufficiently small $V(\bar{y}_1)'$ term with a sufficiently large n_1' then the n_1' is obtained such that

$$V(\bar{y}_1)' + (\bar{y}_1 - \hat{\mu})^2 = MSE_{1+2}$$

that is:

$$n_1' = \frac{s_1^2}{MSE_{1+2} - (\bar{y}_1 - \hat{\mu})^2}$$

This sequence, if my reasoning is correct, yields an estimate of the additional number of type 1 observations, namely $n_1' - n_1$ needed to achieve an estimate as accurate as $\bar{y}_{1+2}$. If the cost of locating and successfully interviewing this number of additional type 1 cases would have been less than what was spent in interviewing the n_2 type 2 cases, then a strategy of interviewing only type 1 cases in other similar surveys is supported.

A second sequence was similarly undertaken to answer the question as to whether it might have been better to limit call back efforts to those which were expended during periods 1 and 2 in which the $n_1 + n_2$ cases, or for consistency with other notation, the n_{1+2} cases were interviewed. The important estimate here is

$$n_{1+2}' = \frac{s_{1+2}^2}{MSE_{1+2+3} - (\bar{y}_{1+2} - \hat{\mu})^2}$$

Hopefully at this point everything will be clear and reasonable except for one major question. What would one use for the estimate $\hat{\mu}$? If, e.g., in the first sequence one used $\bar{y}_{1+2}$ then the bias term in MSE_{1+2} would be 0, which does not make sense. Using $\bar{y}_{1+2+3}$ would be a bit better but would still give a biased, bias. The solution here employed, and the only one I can think of although I would be most interested in help and suggestions on this, was to divide the total number of observations systematically into two subsamples, A and B, obtaining a $\bar{y}_A$ to use in the calculations of all of the statistics based on sample B and a $\bar{y}_B$ to use with A data. This gives two estimates each for n_1' and n_{1+2}' which it would seem reasonable to average. A single variance estimate s_1 used to obtain n_1' from both sample A and B data, was calculated from the combined type 1 cases. Similarly s_{1+2} was calculated using type 1 and 2 cases from samples A and B combined.

Results

The results are not as clear cut as one might hope. We calculated one set of estimates of the number of cases needed with the reduced call-back effort for each of the eleven variables used to make our estimates. Recall that the number of people interviewed in each of the three periods was about 1,600 and that each such group was divided into two equal size subsamples A and B. This means that the n_1' calculation for each sample was based on about 800 cases. We asked, assuming our bias and sampling variance estimates from such a sample were accurate, what would n_1' need to be in order for the MSE based on such a sample size (MSE_1' in the notation used here) to be equal to the MSE_{1+2} estimate based on the approximately 1,600 cases, 800 from each of periods 1 and 2 where one would expect the nonresponse bias to be less. Were our calculations to yield an n_1' somewhat in excess of 1,600, say 2,000, we would think "How reasonable!" Such was the case with the two n_1' estimates using data from the question as to whether or not the physician had graduated from a foreign medical school. Note that, as shown in Table 1 the n_1' based on sample A is 2,044 and the n_1' based on sample B is 1,891 (average 1,968). The two n_{1+2}' estimates based on this variable are also reasonable, perhaps too reasonable or at least too encouraging. The evidence here, assuming 800 in each of the three periods, supports conclusions that our estimates would have been equally accurate had we instead of going to the extra effort expanded during period 3 to interview our 800 physicians, we had interviewed an additional 846 or 877 physicians (average 862) of the type located and interviewed during periods 1 and 2.

REFERENCES

Berk, M.L. (1983,August). Interviewing physicians: The effect of improved response rate. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Deming, W.E. (1953). On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse. Journal of the American Statistical Association, 48, 743-772.

Rao, P.S.R.S. (1966). A study of call-back policies in sample surveys (Technical Report No. 10, Contract Nonr 1866(37), NR-042-097). Washington, D.C.: Office of Naval Research.

Rao, P.S.R.S. (1982). Call-backs: An update. Proceedings of the Section on Survey Research Methods. Papers Presented at the Annual Meetings of the American Statistical Association, 85-87.

Rao, P.S.R.S. (1983). Callbacks, follow-ups, and repeated telephone calls. In W.G. Madow, I. Olkin and D.B. Rubin (Eds.), Incomplete data in sample surveys: Vol. 2, Theory and Bibliographies. New York: Academic Press.

Table 1. Two sets of estimates of n_1', the number of type 1 observations needed to equal MSE_{1+2}, that estimated from n_{1+2} type 1 and 2 observations. Also comparable estimates of n_{1+2}.

	Sample A				Sample B			
	n_{1+2}	n_1'	n_{1+2+3}	n_{1+2}'	n_{1+2}	n_1'	n_{1+2+3}	n_{1+2}'
Income of physician	1208	466	1749	**	1190	*	1716	2335
Days patients wait for an appointment	1543	1275	2314	3694	1550	1323	2325	2798
Physicians -- full time	1632	*	2461	274	1634	*	2460	**
Physicians -- part time	1632	*	2461	2430	1634	*	2464	**
Initial fee	1491	*	2203	**	1483	*	2170	892
Foreign medical student	1621	2044	2440	2446	1620	1891	2441	2477
Age of physician	1575	*	2358	**	1575	7533	2342	390
Whether board certified	1517	*	2271	**	1515	*	2269	3321
Hours per week in patient care	1605	1030	2418	722	1609	*	2422	**
Female	1579	452	2361	**	1589	*	2377	1972
Percent of patients on Medicaid	1446	623	2147	**	1423	541	2129	**

*The bias term $(\bar{y}_1 - \hat{\mu})^2$ is greater than the MSE_{1+2} so, even though the sampling variance term $V(\bar{y})'$ might approach 0, the sum $V(\bar{y})' + (\bar{y}_1 - \hat{\mu})^2$ could never equal MSE_{1+2}

** $(\bar{y}_{1+2} - \hat{\mu})^2 > MSE_{1+2+3}$