COST-EFFICIENCY AND THE NUMBER OF ALLOWABLE CALLBACKS IN THE NATIONAL HEALTH INTERVIEW SURVEY

William D. Kalsbeek, University of North Carolina
Steven L. Botman and James T. Massey, National Center for Health Statistics
William D. Kalsbeek, University of North Carolina at Chapel Hill

1. INTRODUCTION

Many aspects of survey design require trade-offs between statistical efficiency and operational cost. One design option may reduce the mean squared error of estimates but at greater expense than a second option which yields a larger mean squared error. In these instances the designer must often make difficult decisions which account for both statistical and cost implications.

The matter of choosing the maximum number of allowable call attempts to obtain a response from selected households in area samples for personal interview surveys (referred to as the "cutoff" in the sequel) is one design issue which requires these trade-offs. In this instance several cutoff options will be available. However, the larger the cutoff the higher the survey response rate and the smaller the nonresponse bias but the greater the survey cost. Therefore one must decide which cutoff will produce an adequate response rate and yet be affordable.

Our paper addresses the matter of finding the most cost-efficient cutoff for the National Health Interview Survey conducted by the National Center for Health Statistics. The NHIS is an ongoing nationwide sample survey (Massey et al, 1989) which currently does face-to-face interviews in about 48,000 households and where the sample assigned to each quarter of the year is representative of the target population.

Current instructions for NHIS interviewers give high priority to getting a response from selected households and do not prescribe a specific number of call attempts to be made before solicitation attempts cease (U.S. Bureau of the Census, 1985, Section 1.B.2.c). As a result, heroic efforts (i.e., beyond five call attempts) are made to get 4-5 percent of the respondent sample, while the vast majority of the sample is quarried in the first three attempts (Croner, et al., 1985). Although response rates in the 96-98 percent range are realized by this process, one must wonder if the added cost of having virtually no limit on the number of call attempts is justified by the statistical outcome.

In the interest of exploring this issue, NCHS commissioned a study in 1986 to investigate whether the NHIS cutoff could be reduced without appreciably affecting the quality of the survey. Two specific research questions were implicit in this commission. First, how much would the mean squared error of reported NHIS estimates be affected by more severely limiting the number of allowable call attempts? Second, is there some cost-effective cutoff where the response rate would be high enough to control the negative statistical effects (i.e., bias) of nonresponse and yet the direct costs of the field operation would be more moderate?

A study designed to answer these questions is the topic of our paper. We begin by discussing the measures that were used to gauge the statistical implications of the choice of cutoff. We then describe the manner in which the implications of the choice of cutoff on operational cost were measured. Our findings generally reveal that for total population estimates the increase in the mean squared error is only severe when the cutoff is reduced below five attempts. The optimum cutoff varies depending on the measure of interest and the population subgroup to which the estimate applies; however, a cutoff of five or six attempts seems to be a reasonable compromise for all of these situations.

2. METHODS

Data from two sources were used in this study. One is person-level data from the 62,052 respondents to the 1986 NHIS, and the other is a special study of costs on a 10 percent subsample of a portion of the 1987 NHIS that was conducted by the Bureau of the Census in early 1987.

The problem described above is the classic setting for statistical optimization, where a suitable value is found for a parameter with opposite effects on survey costs and the mean square error of estimates. The optimum value of the parameter is found by developing models to represent the statistical and operational impacts of the parameter and then by utilizing some predetermined optimization criterion to arrive at the answer. It is important to note that the statistical impacts measured here will be specific to measures being estimated. Thus, since the statistical impact for different estimates (e.g., average number of physician visits, average number of disability days, etc.) may differ, the optimization solution may likewise differ, making it necessary to somehow pool these findings to arrive at the final recommendation for the parameter.

The approach to optimization that was used in this study is one that might be termed "empirical optimization," where collected data appropriate to the problem are used to obtain direct measures of the statistical and cost impacts of the parameter of interest. This approach avoids the problem of formulating suitable models, although as we shall see there remains the important matter of finding an appropriate way to estimate survey cost and components of the mean squared error as a function of the parameter. The empirical approach seemed most logical in our study, since we had an excellent source of empirical information to obtain the needed figures on statistical efficiency and direct costs in the survey operation.

We let the symbol, c, denote the maximum number of calls attempts that an NHIS interviewer is allowed to make before discontinuing efforts to solicit a response. The formulation of an estimator of a parameter Φ (e.g., the percentage of persons reporting their health as "excellent") would vary somewhat depending on the cutoff. Assuming that the number of selected dwellings is held constant, the final respondent sample size (n_c) and therefore the final adjusted sample weights for the i-th NHIS respondent under a cutoff of c (W_{ci}) will vary by cutoff.

Cutoff-Specific Weights

The approach to computing the W_{ci} in this study followed that which is used in the NHIS Raw selection probabilities were first subject to a weighting class adjustment using the third stage sampling units (i.e., area segments) as adjustment cells. It was at this computational step that the set of weights (W_{ci}) for each cutoff is distinguishable. For a cutoff of c calls this adjustment in the present study was computed as the reciprocal of the product of the call-specific overall NHIS response rate and the relative (to the overall NHIS) rate of response if the maximum number of allowable call attempts was set at c. Each of these rates were truncated at 0.5 as limited attempt at controlling the variation in the final weights. The weighting class adjusted weights for each cutoff were then finally post-stratified to the same 60=15x2x2 age-by-race-by-sex population distribution as used for the 1986 NHIS data that were used in the subsequent analysis.

Thus, along with a different set of respondents for each cutoff, a separate set of weights were used to compute the cutoff-specific estimator of Φ and associated relative measures of its statistical efficacy. Because Φ is a mean or a proportion for all of the parameters that were considered, its estimator under a cutoff rule of c is

$$\varphi_c = \sum_{i=1}^{n_c} W^c_{ci} Y_i / \sum_{i=1}^{n_c} W^c_{ci}. \qquad (1)$$

Relative Bias

The relative bias of φ_c, Rel-Bias(φ_c) $\equiv$Bias(φ_c)/Φ, where Bias(φ_c)=E(φ_c)-Φ. Assuming that the amount of remaining bias due to nonresponse for the estimate of Φ obtained from the full NHIS sample (φ_+) is small, then a reasonable estimator of Bias(φ_c) would be bias(φ_c)=(φ_c-φ_+), from which relative bias can be estimated as.

$$\text{rel-bias}(\varphi_c) = \text{bias}(\varphi_c)/\varphi_+. \qquad (2)$$

This measure can also be viewed as the change in the estimate of Φ at a cutoff of c, relative to the same estimate as obtained under the existing callback strategy.

Relative Change in Standard Error

Another measure of the statistical implication of setting the cutoff at c is the relative change in the standard error of estimates. This second measure serves to indicate the relative decrease (n_c diminishes with increased c) in the precision of NHIS estimates as, or equivalently the relative increase in the width of interval estimates as the number of allowable call attempts is modified from current practice. This measure is defined as,

$$D_{SE}(\varphi_c) = [SE(\varphi_c)-SE(\varphi_+)]/SE(\varphi_+),$$

where SE(φ_c) is the standard error of φ_c of the estimate of Φ under a cutoff of c, and SE(φ_+) is the comparable standard error from the full NHIS sample.

Using the statistical program SESUDAAN (Shah, 1981) to generate the estimates, se(φ_c) and se(φ_+), of SE(φ_c) and SE(φ_+), respectively, we estimate $D_{SE}(\varphi_c)$ as,

$$d_{SE}(\varphi_c) = [se(\varphi_c)-se(\varphi_+)]/se(\varphi_+). \qquad (3)$$

Relative Root Mean Squared Error

A third measure of statistical impact is needed to assess the overall statistical effect of the cutoff strategy. The mean squared error of φ_c is by definition,

$$MSE(\varphi_c) = Var(\varphi_c) + Bias^2(\varphi_c),$$

which can be estimated as

$$mse(\varphi_c) = se^2(\varphi_c) + bias^2(\varphi_c).$$

The measure of statistical efficacy we ultimately used was the relative root mean squared error,

$$RMSE(\varphi_c) \equiv [MSE(\varphi_c)]^{1/2}/\Phi,$$

which we estimated as,

$$rmse(\varphi_c) = [mse(\varphi_c)]^{1/2}/\varphi_+. \qquad (4)$$

We are reminded by using Var(φ_c) that we presume that alteration of the cutoff would leave unaffected the number of selected dwellings in the NHIS sample but would affect the number of responding households.

Direct Costs

Unlike most cost-efficiency studies within a survey design context, where one must speculate on the costs based on a mathematical model to which various estimated parameters of unit cost are applied, cutoff-specific costs for the present study could be estimated directly from data gathered by Bureau of the Census field interviewers for a random subset of the NHIS household sample.

Households in the NHIS costs substudy, that was designed to assess total direct costs of the NHIS field operation, were those falling in a roughly 10 percent subsample of (63) segments from the second quarter sample of the 1987 NHIS. Proportionate stratified selection of segments from five Census' "production strata" (i.e., basically reflecting population density) was done to improve the precision of estimated costs.

Not all of the selected segments could be used in the study. One had been merged with another segment, two had to be dropped from the study for practical reasons, and the data on two more were lost. This left 58 segments from which costs data were obtained.

Several units of significance must be defined to describe data collection for the costs study and the method by which costs were estimated here. An interviewer assignment area, usually linked to one interviewer except in the cases where shuffling of field staff was required to expedite the completion of interviewing, consisted of one or more sample segments which, in turn, consisted of several dwellings serving as sampling units that were subject to one or more call attempts in obtaining a response. To complete the field operation in each assignment one or more trips were made, some as part of one or more GTR travel events.

A detailed accounting of the outcome of solicitation efforts and associated costs was obtained for each assignment by means of a specially devised but simple administrative form for the interviewers to complete. Costs accounted for on this form were those for interviewer time, mileage, and other expenses related to data collection in the assignment. Specifically excluded from the direct costs were the time and expenses for interviewing supervisors and costs for ancillary interviewer activities such as training, updating of lists of dwellings used for sampling, manual data editing, questionnaire disposition, and home activities like calling to make or confirm appointments. Since rates of monetary reimbursement for project work varied somewhat among interviewers, time and expenses to dollar costs were uniformly converted among interviewers, using conversion factors of \$0.131/minute and \$0.205/mile, respectively.

The estimated total direct cost of operations (including the cost of all necessary trips and GTR travel) for work up through and including c visits to each sample unit was estimated as

$$cost_c = K\,[\sum_{i=1}^{m_c} C_{ci} + \sum_{j=1}^{58} C_{cj}] \qquad (5)$$

where m_c is the final sample size under a cutoff of c attempts in the costs substudy, K = 43.45 = 10x(63/58)x4 is an inflation factor to account for the initial sampling rate of segments, segment attrition, and the fact that the study was conducted on a one-fourth sample of the 1987 NHIS, C_{ci} is the cutoff-specific cost of interviewing time, and C_{cj} is the estimated travel cost for work in the assignment, if the cutoff had been at c calls.

The key to obtaining C_{cj} was our ability to reconstruct details for the sequence of events which led to the completion of work in each assignment area. C_{cj} included two components. One was the directly reported salary and expenses for trips to the assignment area if, based on the history of call attempts in that area, the trip would have been necessary. So, for example, if the final trip to a segment was only to do the fourth call at one household and the fifth at another, that trip would not have been figured into values of C_{cj}, where $c \leq 3$. The second component of C_{cj} was an estimate of the per diem and other travel costs for travel required to the assignment, given the cutoff. Since these costs had not specifically been linked to each trip to the assignment area, we had to estimate these costs for each trip by prorating the aggregate of these miscellaneous costs according to the relative amount of interviewing time that occurred during the trip. For example, if 20 percent of the total interviewing minutes for the assignment took place on the third trip, then for whatever cutoff that trip would have been needed the total miscellaneous costs for that assignment, times 0.20, would have been figured into C_{cj}.

Domain-Specific Costs

It must be mentioned that while estimating the statistical efficacy of estimates for various population subgroups (e.g., by age, race, sex, education) was possible, it was not possible to effectively estimate domain-specific costs since the necessary demographic information was not collected in the costs study, nor could the costs data be linked to demographic information collected in the 1987 NHIS. Moreover, even if matching could have been done, there would have still remained the difficulty of attributing certain of the costs to a specific subgroup (e.g., travel time and expenses on individual trips to interview persons from more than one subgroup).

Cost-Efficiency

One set of $cost_c$ (c=1,2,...,8) were produced, with a ninth value, $cost_+$, to represent the estimated cost of NHIS data gathering under the present system. The measure of cost-efficiency, given a cutoff at c calls, that was used in this study was computed as,

$$CEFF_c = \frac{1}{cost(\varphi_c)\,[mse(\varphi_c)]^{1/2}} \qquad (6)$$

To eliminate the effect of variation in scale for the $CEFF_c$ produced for differing Φ, the level of $CEFF_c$, relative to the maximum of the $CEFF_c$ among the 9 candidate cutoffs, computed as,

$$RCEFF_c = CEFF_c / \max_c(CEFF_c) \qquad (7)$$

3. FINDINGS

Measures of the statistical and cost implications of choosing a particular cutoff, as well as the most cost-efficient choice of c, are presented for several measures from which estimates are commonly reported in NHIS publications. These include:

(1) HOSPEPIS = The average number of short-stay hospital visits in the last six months;
(2) HSTATUS = The percentage reporting their health status as "excellent";
(3) MDVISIT = The average number of physician visits in the last two weeks; and
(4) RESTRACT = The average number of restricted activity days in the last two weeks.

Statistical implications on estimates for these measures were assessed for total population estimates as well as for various subpopulations defined by gender, age, and race.

Table 1 contains some of our findings by cutoff for the full NHIS sample as well as for the aggregate of various demographic subgroups. It is interesting to note that direct costs, relative to current practice, under the eight alternative cutoff options to current practice is very similar to comparable relative response rates. This implies that the values of $cost_c$ and n_c are highly correlated.

Measures of percent rel-bias(φ_c) are also presented in Table 1 for four common NHIS estimates. As expected, the level of bias decreases as the cutoff in the number of allowable calls (c) increases since response rates improve with more allowable call attempts. The direction of the rel-bias for these estimates indicates that NHIS nonrespondents (relative to respondents) under these cutoffs would tend to: (1) have fewer hospital visits, (2) be more likely to report being in excellent health, (3) have fewer doctor visits, and (4) be less restricted in their daily activities. These findings are supported by the contents of some other (unreported) analyses we did which show that to varying degrees late respondents to the NHIS tend to be: (1) in the relatively healthy 17-44 age group, (2) in the middle to upper income groups, (3) better educated, and (4) living in single person households.

The pattern in the percent change in the standard error of estimates in Table 1 is similar for all estimates. Values of $d_{SE}(\varphi_c)$ drop off rapidly after an allowable cutoff of three or four calls. This pattern parallels the pattern of increased response rate and is thereby largely due to the larger respondent sample sizes, n_c, as c increases.

Values of the relative cost-efficiency of each alternative cutoff are also presented for the four NHIS estimates in Table 1. Glancing across these rows reveals how cost-efficient each cutoff option would be for the estimate, relative to the empirically optimum option (i.e., with $RCEFF_c = 1.00$).

Except for MDVISIT, where the optimum choice is at a single call attempt, several values of $RCEFF_c$ fell between 0.90 and the 1.00 for the most cost-effective cutoff (c^*), indicating that the choice of the c^* is not always clear-cut and that other choices may be made without an appreciable drop in efficiency and costs. This "flatness" in cost-efficiency tended to occur for larger values of c^* where there were smaller differentials in both mse(φ_c) and cost(φ_c).

The most cost-effective cutoff is presented for the total population as well as various demographic subgroups in Table 2. A few other differentials, or lack thereof, on c^* were noteworthy here. First, there were few differences by gender, except for HOSPEPIS although the optima here were in relatively flat areas of cost-efficiency where little real difference in $CEFF_c$ would have been realized whether, for either subgroup, c^*=4 or c^*=8 been adopted. Second, difference for racial groupings did exist but for larger values of c^* where several values of $CEFF_c$ were relatively equal.

The instances (for MDVISIT in Table 1 and for several subgroups in Table 2) where c^*=1 were not characterized by the flatness mentioned above but rather by relatively equal but notably smaller values of $RCEFF_c$ for c>1. They tended to occur when the bias due to nonresponse was relatively small after only one call attempt and when cost is relatively small as well.

We also determined (but did not have room to present here) differentials on rel-bias(φ_c) and the relative root mean squared error (rmse(φ_c)) by age. Regarding measures of relative bias among alternative cutoffs, the patterns observed among age groups generally mimic the pattern of the total population estimate but are either more severe or less severe. Isolated middle age groups (35-54) were especially likely to exhibit extreme patterns with relatively severe bias occurring for smaller values of c. For the most part, the direction of bias in these age groups followed the direction of bias for the corresponding estimate for the total population. Patterns for rmse(φ_c) among these same age groups once again largely mirrored the pattern of rmse(φ_c) as seen for total population estimates, and as with rel-bias(φ_c) there were some age groups with amplified patterns. While these extreme patterns can once again be found in the 35-54 age groups, there appear elsewhere as well. For example, they also appear in the older age groups for HSTATUS and MDSVISIT, and for HOSPEPIS for those <6 years.

Average values of RCEFF by cutoff option for NHIS total population and demographic subgroups are presented at the bottom of Table 1. One minus these entries indicates the average relative loss in cost-efficiency one would realize for NHIS estimates in the aggregate under the nine cutoff options. We see that for total population estimates the smallest average loss (12 percent) occurs when the cutoff is set at six attempts. It is worth noting, however, that average losses for any cutoff from five up to current practice have similar losses at the numerical optimum. As noted previously, the smallest average loss for subgroups estimates is for a cutoff after a single call, where the average loss is 13 percent. This finding is the result of the relatively large number of subgroups (especially those for MDVISITS) where c^*=1. Assuming then that this result may be an artifact of our choice of NHIS-reported health characteristics and that prospects of an NHIS response rate of less than 40 percent (if c=1 were chosen) are unappealing, then we must look at the entries for c>1 for our recommended choice for c. Doing so, we see once again that any cutoff at five or greater would result in roughly equivalent average losses in RCEFF.

Clearly more thought must be given to the matter of choosing the best cutoff for NHIS. For example, perhaps other estimates should be considered and the relative importance of these estimates for the total population and demographic subgroups prioritized through some sort of weighting scheme. In any event the multi-purpose nature of NHIS and all surveys demands careful consideration in arriving at a final decision on the optimum number of allowable call attempts.

REFERENCES

Census Bureau (1985). National Health Interview Survey :11.

Croner, C.M., Williams P.D. and Hsiung S. (1985). "Callback Response in the National Health Interview Survey," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 164-169.

Massey, J.T., Moore, T.F., Tadros, W., and Parsons, V.L. (1989) "Sample Design and Estimation for the National Health Interview Survey, 1985-1994," Vital and Health Statistics Series, Series 2, No. 110.

Shah, B.V. (1981) "SESUDAAN: Standard Errors for Survey Data Analysis," Research Triangle Institute, Box 12194, Research Triangle Park, NC 27709.

Table 1. Cost and Statistical Implications for Selected NHIS Estimates by Number of Allowable Call Attempts to Households.

	Number of Allowable Call Attempts								
	1	2	3	4	5	6	7	8	Current Practice
Direct cost relative to current practice	39.0	63.8	79.3	87.4	91.9	95.1	97.4	98.3	100.0
Response rate relative to current practice	39.6	65.0	80.2	88.2	93.2	96.8	98.0	98.8	100.0
Hospital visits per person:									
Percent relative bias	10.4	5.4	2.6	0.8	0.8	0.1	0.1	-0.0+	0.0
Percent change in standard error (d_{SE})	73.7	34.8	17.5	6.7	2.8	-1.3	-1.2	-2.7	0.0
Relative cost efficiency (RCEFF)	0.40	0.44	0.63	0.92	0.90	1.00	0.97	0.98	0.94
Percent reporting excellent health:									
Percent relative bias	-3.9	-2.3	-1.2	-0.1	0.3	0.0+	0.0+	0.1	0.0
Percent change in standard error (d_{SE})	37.3	16.6	-1.5	1.8	-1.6	-3.4	0.8	2.1	0.0
Relative cost efficiency (RCEFF)	0.62	0.60	0.79	1.00	0.96	0.98	0.91	0.89	0.90
Doctor visits per person:									
Percent relative bias	2.0	2.8	2.3	1.9	1.2	0.6	0.6	0.4	0.0
Percent change in standard error (d_{SE})	7.4	18.3	15.7	9.0	3.3	2.5	4.6	0.6	0.0
Relative cost efficiency (RCEFF)	1.00	0.49	0.44	0.46	0.52	0.57	0.55	0.57	0.58
Restricted activity days per person:									
Percent relative bias	7.8	5.4	3.4	2.5	1.2	0.9	0.8	0.3	0.0
Percent change in standard error (d_{SE})	37.7	23.2	17.2	9.6	2.6	1.1	4.1	-0.1	0.0
Relative cost efficiency (RCEFF)	0.69	0.58	0.66	0.72	0.92	0.96	0.92	1.00	0.99
Average RCEFF among estimates:									
Overall population*	0.68	0.53	0.63	0.78	0.83	0.88	0.84	0.86	0.85
Population subgroups**	0.87	0.72	0.73	0.76	0.78	0.80	0.79	0.80	0.79

Note: 0.0+ indicates that the value is somewhere between zero and 0.05.

* Averaged over the four health characteristics regularly estimated from NHIS data and for which statistical measures are produced earlier in this table.

** Averaged over 164 values obtained from 41 demographic subgroups defined by gender, family income, age, race, education of the head of household, and family size for each of the same four health characteristics regularly estimated from NHIS data.

Table 2. Optimum Cutoff for Selected NHIS Estimates of Health Characteristics by Population Subgroup

	Health Characteristic			
Subgroup	Number of Short Stay Hospital Episodes in Past 12 Months	Health Status Excellent	Number of Physician Visits in Last 2 Weeks	Number of Restricted Activity Days in Past 2 Weeks
Total	6	4	1	8
Male	6	4	1	8
Female	4	4	1	9
Under 6	1	1	1	1
6-16	1	1	6	1
17-24	2	4	8	1
25-34	8	1	1	1
35-44	6	5	1	8
45-54	3	3	1	6
55-64	1	4	1	8
65-74	1	2	6	2
75+	1	1	1	1
White	4	4	1	8
Black	7	3	1	1
Other	6	1	9	3