Journal of Official Statistics
Vol. 10. No. 2, 1994, pp. 133–152

Cost-Efficiency and the Number of Allowable Call Attempts in the National Health Interview Survey

William D. Kalsbeek,[1] Steven L. Botman,[2] James T. Massey,[2] and Pao-Wen Liu[1]

Abstract: An important design decision for interview surveys is the number of call attempts allowed in obtaining a response from sample members. Considerations of survey cost favor a small number of allowable attempts, while the desirability of a high response rate points to a large number. An empirical assessment of the statistical efficiency for set cost among call attempt options sheds some light on this issue as it pertains to the U.S. National Health Interview Survey. Cost data were obtained from a special assessment on a subsample of the NHIS, and efficiency is measured by the estimated mean squared error of various estimates from the NHIS sample. While the current practice of allowing up to 15 call attempts is found to be the best overall, the cost-efficiency of cutoffs with as few as six attempts is nearly as good.

Key words: Nonresponse; callbacks; survey cost; design optimization.

1. Introduction

Many decisions in survey design require a trade-off between statistical efficiency and operational cost. One design option may yield a smaller mean squared error of estimates, but at greater expense than another. Choosing the number of allowable attempts to obtain a response (i.e., the problem of identifying the best "cutoff" for response solicitation) is one design issue that requires a trade-off involving the sample response rate and survey cost. The larger the cutoff, with the higher response rate and smaller nonresponse bias it produces, the greater the survey cost. Therefore one must decide which cutoff will produce an adequate response rate and still be affordable.

Studies which examine the role of the allowable number of solicitation attempts (or call attempts) to obtain a response in sample surveys but which ignore the effect on cost have been published. Some of the early ones, where respondent outcomes were differentiated by call attempt, have been reviewed in survey methods texts (e.g., Cochran 1977; Kish 1965; and Zarkovich 1963). The profiles of attempt-specific response rates and estimates presented in these studies clearly demonstrate the statistical effect of the cutoff decision. Typical

[1] Survey Research Unit, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-2400, U.S.A.
[2] Office of Research and Methodology, National Center for Health Statistics, Hyattsville, MD 20782, U.S.A.
Acknowledgements: We wish to recognize Steve Mann, Tom Moore, and Bob Wilson of the U.S. Bureau of the Census for their efforts in producing the NHIS cost data we used for our research. We also would like to thank James Locklear for the excellent statistical programming support he provided to this project. Finally, we gratefully acknowledge the many helpful suggestions made by the Associate Editor and two Referees.

of this approach is a study by Rao (1983) in which these kinds of comparative profiles are presented for three surveys, two using mail data collection and one done by face-to-face interview. Rao (1982) has also presented an overview of research on how cutoff strategies affect nonresponse and on setting the number of allowable attempts.

A number of purely empirical studies have used estimated statistical effects to recommend a suitable cutoff strategy for call attempts in interview surveys. For example, Croner, Williams, and Hsiung (1985) have used measures of the mean squared error of estimates for the National Health Interview Survey (NHIS) to suggest a cutoff. In another study, Fitch (1985) chose a cutoff option by estimating the change in the mean squared error brought about by reducing the number of call attempts and increasing the overall sample size.

The search for a cutoff strategy that seeks a suitable compromise between the countervailing effects of survey error and cost has also been given some attention, although the optimization framework in these studies has varied. For example, Drew and Fuller (1981) and Drew and Ray (1984) found an optimum calling strategy by using a bias model that assumes response propensity to be a function of a discrete survey variable. Rao (1982), on the other hand, reviewed several approaches for optimizing a calling strategy which involves selecting subsamples of nonrespondents at successive call attempts. More recently, Groves (1989, sec. 4.5.3) has suggested an approach in which components of error and cost models are estimated from a random subset of the survey sample and then used to find a cutoff option that minimizes mean squared error for set cost among options.

The effect of the calling strategy on survey costs is partly determined by the sample's urban–rural mix. Interviewers in urban areas have greater difficulty in securing a response through added call attempts than do their rural counterparts (Dunkelberg and Day 1973), and national samples are often highly urban. Thus, the conclusion by Steeh (1981) that the degree of urbanization is a major exogenous factor in predicting nonresponse bodes poorly for survey costs when more call attempts are made to improve response rates. Even in rural areas, survey costs can be substantially increased by allowing multiple callbacks, since return calls may involve considerable travel expense.

This paper addresses the matter of finding the most cost-efficient cutoff strategy for the NHIS, conducted for the National Center for Health Statistics by the U.S. Bureau of the Census (Massey, Moore, Tadros, and Parsons 1989). The NHIS is an ongoing nationwide sample survey in which face-to-face interviews typically have been done in about 48 thousand residential dwellings each year and where the sample assigned to each quarter of the year is representative of the civilian noninstitutionalized U.S. population. The NHIS includes basic health and demographic data as well as data on current health topics added as a supplement to each year's questionnaire.

Training instructions for the NHIS stress high response rates. Interviewers are allowed up to 15 attempts before discontinuing efforts to obtain a response from sample households (U.S. Bureau of the Census 1985, sec. 1.B.2.c). As a result, extensive efforts (i.e., more than five call attempts) are made to get 4%–5% of the respondent sample, while the vast majority of the sample is queried in the first three attempts (Croner et al. 1985). Although response rates in the 95%–98% range are realized each year by the NHIS, this paper addresses whether the added cost of this cutoff strategy is justified by the resulting data quality.

To explore this issue, NCHS began a research study in 1986 to investigate whether the number of allowable call attempts could be reduced without appreciably affecting the survey's data quality. Two specific research questions were identified for that study. First, how much would the mean squared error of reported NHIS estimates be adversely affected by limiting the number of allowable callbacks? Second, is there some cost-effective cutoff where the response rate would be high enough to yield moderate levels of nonresponse bias and where data collection costs are more moderate than those required under current practice in the NHIS?

We first present the statistical measures used to gauge the relative efficiency among cutoff options for various subpopulation estimates of means and proportions that are annually reported by NCHS from the NHIS. This is followed by a description of the approach that was used to estimate the operational cost for each cutoff strategy. Cost and statistical implications are then tied together to identify, for a set of NHIS estimates, the options that produce the lowest mean squared error for set cost. Finally, the limitations and implications of our findings are discussed.

2. Estimation and Statistical Measures

The statistical measures used for this study were obtained for total population and subpopulation estimates based on data collected from the 62,052 persons who live in households that completed the 1986 NHIS interview. We use the symbol, c, to denote a designated "cutoff"; i.e., the maximum number of call attempts that an NHIS interviewer would be allowed to make in soliciting participation from selected dwellings. The cutoff in current practice with the NHIS is $c = 15$.

Since the call attempt on which the response was obtained was known for all NHIS respondents, a sample of individuals who would have been respondents under a cutoff at c call attempts could be identified by the subset of NHIS respondents whose data were obtained on any attempt up to and including the cth call attempt. Cutoff-specific estimates (ϕ_c) of a parameter (Φ) were obtained from these respondent subsets, although separate sampling weights (W_{ci}) were needed to compensate for the varying patterns of nonresponse among cutoff options.

Several statistical measures associated with ϕ_c were computed to gauge the size of various statistical effects of each cutoff option. Some of the measures were relative measures, computed as the cutoff-specific measure divided by some referent value. In addition to using the best available estimate of Φ, maximum and minimum values of the measure among options, as well as the measure associated with the current practice, were used as the referent value. Many of these relative measures were developed to depict how close each cutoff-specific measure is to the best or worst measure among cutoff options.

Statistical measures were obtained for fifteen person-level health indicators that are annually estimated by population subgroup from the NHIS. Ten of the estimates were for proportions associated with the possible responses to the following categorical variables:

1. The person's current health status: excellent, very good, good, fair, poor, or unknown; and

2. The person's current limitation status: unable to perform major activity, limited in the kind/amount of major activity, limited in the kind/amount of other activities, or not limited.

Five of the estimates were for means of the following discrete variables defined for persons responding to the NHIS:

1. The number of current medical conditions;
2. The number of days spent in short-stay hospitals during the last six months;
3. The number of short-stay hospital visits in the last six months;
4. The number of physician visits in the last two weeks; and
5. The number of restricted activity days in the last two weeks.

The statistical measures were for estimates obtained for the total civilian noninstitutionalized U.S. population as well as for estimates applied to 41 subpopulations defined by univariate groupings on sex, age, race, education, family income, and family size with 2, 9, 3, 8, 10, and 9 categories, respectively.

2.1. *Cutoff-specific sampling weights and weighted estimates of Φ*

The production of subpopulation estimates for each cutoff option followed standard NHIS practice (see Massey et al. 1989). Sampling weights for the subset of n_c respondents were post-stratified to the joint U.S. population distribution by age (15 categories), race (2 categories), and sex (2 categories), and the usual weighted ratio-type estimator

$$\phi_c = \sum_{i=1}^{n_c} W_{ci} Z_i Y_i \Big/ \sum_{i=1}^{n_c} W_{ci} Z_i$$

was used to produce cutoff-specific subpopulation estimates of Φ, where $Z_i = 1$ if the ith respondent is a member of the subpopulation, and $Z_i = 0$ if otherwise. Here Y_i is the measure of interest for the ith respondent. When used to estimate a population mean, Y_i is a discrete measure (e.g., number of current medical conditions). For estimates of proportions, the value is dichotomous depending on whether $(Y_i = 1)$ or not $(Y_i = 0)$. The respondent chooses the category of interest, e.g., current health status is reported as "excellent."

2.2. *Variance and design effect of ϕ_c*

A measure of the sampling error of estimates resulting from having set the cutoff at c is the variance of ϕ_c, $\mathrm{Var}(\phi_c)$. The estimate, ϕ_c, and an estimate of its variance, $\mathrm{var}(\phi_c)$, were produced directly from NHIS respondent data using the statistical program SESUDAAN (Shah 1981), which follows the Taylor series approach to variance estimation for data generated from complex sample designs.

The overall design effect associated with ϕ_c, defined as the variance of ϕ_c from the actual NHIS sample design divided by the variance one would expect if an NHIS sample of the same size were chosen by simple random sampling, can be written as

$$\mathrm{Deff}_{oc} = \mathrm{Var}(\phi_c)/(S^2/n_c)$$

where n_c is the NHIS respondent sample size for a cutoff at c calls, and S^2 is the variance of the population measure tied to Φ (Kish 1965). Deff_{oc} can thus be estimated as

$$\mathrm{deff}_{oc} = \mathrm{var}(\phi_c)/(s_c^2/n_c) \qquad (1)$$

where s_c^2 is a weighted estimate of the element variance, S^2, based on data from the same set of n_c respondents that is used to produce ϕ_c.

2.3. *Bias and relative bias of ϕ_c*

Another summary measure of survey error is the relative bias of ϕ_c.

$$\text{Rel-Bias}(\phi_c) = \text{Bias}(\phi_c)/\Phi$$

where the bias of ϕ_c is defined as

$$\text{Bias}(\phi_c) = E(\phi_c) - \Phi.$$

The size of Bias(ϕ_c) is expected to be inversely related to the response rate for a cutoff at c attempts (Rate$_c$), defined as the proportion of eligible population members who would respond within c attempts, if they were chosen in the sample.

Since under current practice the overall response rate for the 1986 NHIS was very high (96.5%), we assume that the amount of remaining bias due to nonresponse for the estimate of Φ under the current cutoff policy (ϕ_{15}) is negligible. Thus, a plausible estimator of Bias(ϕ_c) from NHIS data is

$$\text{bias}(\phi_c) = \phi_c - \phi_{15}. \tag{2}$$

And the relative bias can be estimated as

$$\text{rel-bias}(\phi_c) = \text{bias}(\phi_c)/\phi_{15}. \tag{3}$$

The rel-bias(ϕ_c) measure can therefore also be viewed as the change in the estimate of Φ for a change in cutoff strategy, relative to the same estimate based on current practice. Rate$_c$ can also be estimated as

$$\text{rate}_c = 0.965(n_c/n_{15}) \tag{4}$$

where 0.965 was the overall response rate for the NHIS at the time of the study.

3. Cost Measures

The direct operational cost of each cutoff option was estimated from a special study of NHIS data collection costs. The study was performed on a stratified subsample of dwellings from the NHIS sample for the second quarter of 1987. The data in this cost study were gathered by Bureau of Census field interviewers using time and expense sheets that had been designed to provide a detailed breakdown of costs associated with each phase of data collection. Reported cutoff-specific costs were obtained by aggregating those expenses that would have been incurred by the NHIS only through the number of interviewer call attempts allowed under the cutoff.

3.1. *NHIS terminology*

Several terms linked to features of the data collection process in the NHIS are used in later sections. An interviewer *assignment* area consists of one or more selected areas or permit *segments* (second stage sampling units) in the NHIS multistage sampling design. Each sample segment, in turn, consists of several *dwellings*, which are the units sampled within the segment. One or more call attempts are made to obtain a response from each sample dwelling. Call attempts on a dwelling are usually made by the same interviewer except when a reassignment of field staff is needed to expedite interviewing. To complete the field operation in each assignment, one or more trips are made to each segment, usually from the interviewer's home. When assignment areas are far from the interviewer's home, a Government Travel Request (GTR) travel event may be authorized for extended travel to the assignment. As part of such an authorization, a *per diem* may be given and other related expenses covered including airline tickets, train tickets, etc.

3.2. *Cost study design*

The cost study sample was chosen by selecting an approximately proportionate stratified sample of 63 interviewer assignment areas (out of a total of 781) from five census "production strata" that differ according to population density. Not all of the selected assignment areas could be used in the study. One had been merged with another, two had to be dropped for practical reasons, and the data on two more were lost. This left the entire sample of dwellings in 58 areas for which call-specific costs of solicitation and interviewing were obtained.

The cost study yielded an accounting of the outcome of solicitation efforts and associated costs for each assignment by means of a specially devised, but simple, administrative form that was completed by a designated Census Bureau "field representative," whose only role was to complete the form. Costs accounted for on the form were for interviewer time, mileage, and other expenses related to data collection activities in the assignment area. Not included were costs for the time and expenses of interviewing supervisors and for interviewer time for training, listing and updating, and field sampling. Costs for other ancillary interviewer activities like recording outcome information for call attempts, logging final nonrespondents, telephone callbacks to complete interviews, and questionnaire field editing were taken into account by the cost estimates. Because rates of monetary reimbursement for project work varied slightly among interviewers, uniform conversion factors of \$0.131/minute and \$0.205/mile were used for salary and mileage, respectively.

3.3. *Estimates of cutoff-specific cost*

The annual cost of interviewer-related data collection in the NHIS for a cutoff at c attempts was estimated as

$$\text{cost}_c = \sum_{h=1}^{5} W_h \left\{ \sum_{i=1}^{a_h} \left[F_{hi} + T_{hic} + \sum_{j=1}^{d_{hi}} (S_{hijc} + U_{hijc}) \right] \right\} \tag{5}$$

where for the hth production stratum: $W_h = 4A_h/a_h$ is an inflation factor accounting for sampling and attrition in the cost study; A_h is the number of assignment areas on the sampling frame for the cost study; a_h is the number of sample assignments for which cost study data were obtained; d_{hi} is the number of NHIS sample dwellings in the ith assignment chosen for the cost study; F_{hi} is the survey cost in the ith assignment that is not influenced by cutoff option (i.e., for GTR transportation fares, salary, mileage, and other expenses for travel to and from the assignment); S_{hijc} is the interviewer salary for activities tied to successful interview attempts through the cth call attempt for the jth dwelling in the ith assignment (i.e., for time spent in conducting interviews, logging the outcome of those attempts, and manual editing of completed interviews); U_{hijc} is the interviewer salary for activities tied to unsuccessful interviewing attempts through the cth call attempt for the jth dwelling in the ith assignment (i.e., for time spent in making unsuccessful attempts to complete an interview, for logging the outcome of those attempts, for final outcome logging of nonrespondents); and T_{hic} is the cost of travel to and from the segment as well as all other survey activities in the ith assignment that depend on cutoff options but are not wholly considered to be costs of successful or unsuccessful attempts (i.e., for interviewer salary and mileage on all trips to the segment that were needed to complete work in the assignment under a cutoff at c attempts, as well as apportioned per diem expenses for any extended travel to the assignment that might have been required).

Information on interviewer salary to determine S_{hijc} and U_{hijc} was usually available from cost study data. The two exceptions were final outcome logging for nonresponding dwellings and manual editing of completed questionnaires. Operational data from Field Division staff at the Census Bureau were used to estimate unit costs for these two cost components. The cost of nonrespondent outcome logging for a given cutoff option was thereby estimated as the number of final respondents times $0.66 = (5 logging minutes per final nonrespondent) × ($0.131 per salary minute), while the estimated cost of manually editing completed questionnaires was computed as the number of respondents times $5.72 = (43.64 editing minutes per respondent) × ($0.131 per salary minute).

The key to identifying the set of trips linked to T_{hic} was to reconstruct the sequence of events that led to the completion of work in the NHIS sample segments in each assignment area. And, T_{hic} was computed as the sum of two components. One was the directly reported mileage and other trip costs for travel to and from the sample segment if, based on the history of call attempts at dwellings in the segment, the trip would have been necessary. So, for example, if the sole purpose of the final trip to a segment was to do the fourth call at one dwelling and the fifth at another, that trip would not have figured into the computation of T_{hic} for $c \leq 3$.

The second component of T_{hic} was an estimate of per diem expenses for travel to the assignment, given the cutoff. The portion of per diem cost allocated to a trip to the segment was directly proportional to the total cost of the trip. This allocation of assignment-level costs was needed because the length of time in the assignment (for which per diem expenses are paid) will vary depending on the extent of travel required to sample segments in the assignment area. For example, if 20% of the total cost of interviewer travel to segments in an assignment area was attributable to a given trip and that trip was needed for a cutoff at c attempts, then the per diem contribution to T_{hic} from that trip was computed as the total assignment-level per diem times 0.20.

3.4. *Subpopulation-specific costs*

Although calculating the statistical efficiency of estimates for various subpopulations (e.g., by age, race, and education) was possible, it was impractical to estimate subpopulation-specific costs because demographic information was not available for the households in the cost study. Moreover, even if matching could have been done, not all of the costs (e.g., for travel time and expenses on individual trips to interview persons from more than one subpopulation) could be attributed to a specific subpopulation.

4. Cost-Efficiency Measures

Following Groves (1989; sec. 4.5), the relative cost-efficiency of each cutoff option for the NHIS was assessed for several subpopulation estimates of Φ. Simple models for the cost and mean squared error associated with ϕ_c were necessary. With cost set at some level for all cutoff options, the respondent sample size that would produce that cost was determined for each cutoff. The set-cost sample size for the cutoff was then applied to the mean squared error model to produce a cutoff-specific measure of statistical efficiency. The largest efficiency (i.e., smallest mean squared error) among cutoff options thus indicated the greatest cost-efficiency.

4.1. Cost model

The model for our cost-efficiency comparisons among cutoff options isolates components of data collection cost for responding and nonresponding dwellings. It also presumes that the field staff configuration cannot be easily changed, thus requiring that the number of sample PSUs for the NHIS remain constant among options. The total cost for the cth cutoff is thereby expressed as

$$\begin{aligned} \mathrm{Cost}_c &= F + T_c + S_c + U_c \\ &= F + T_c + d_{c1}\mu_{c1} + d_{c0}\mu_{c0} \qquad (6) \end{aligned}$$

where F is the fixed cost assignment-level costs (aggregated from F_{hi} in Equation 5); T_c is the travel and other costs affected by c (aggregated from T_{hic} in Equation 5); $S_c = d_{c1}\mu_{c1}$ is the cost of all successful call attempts (aggregated from S_{hijc} in Equation 5); $U_c = d_{c0}\mu_{c0}$ is the cost of all unsuccessful call attempts (aggregated from U_{hijc} in Equation 5); d_{c1} and d_{c0} are, respectively, the cutoff-specific number of responding and nonresponding sample dwellings; and μ_{c1} and μ_{c0} are, respectively, the average per respondent cost for successful call attempts and the average per nonrespondent cost for unsuccessful attempts for the cth option. Here, F is assumed to be the same for all cutoff options since a constant PSU sample size implies the same set of assignment areas and thus fixes the aggregate of assignment travel costs that are unaffected by option. Since only total population estimates of cost could be derived from the cost study, μ_{c1} and μ_{c0} for the total population were also assumed for each subpopulation.

4.2. Variance model

We assume that the variance of ϕ_c can be modelled as

$$\begin{aligned} \mathrm{Var}(\phi_c) &= \mathrm{Deff}_{oc}S^2/n_c \\ &= \mathrm{Deff}_{Wc}\,\mathrm{Deff}_{Ic}S^2/n_c \qquad (7) \end{aligned}$$

where n_c is the respondent sample size (in persons) for the NHIS estimates that were considered, Deff_{oc} is the overall design effect (i.e., amount of increased variance) of ϕ_c, which can be estimated by programs like SESUDAAN, Deff_{Wc} is the multiplicative increase in the variance of ϕ_c due to variation in the sampling weights that are used to produce ϕ_c, and

$$\mathrm{Deff}_{Ic} = 1 + \mathrm{Roh}_{Ic}(n_c/m - 1) \qquad (8)$$

is the design effect due to cluster sampling, where Roh_{Ic} is a measure of intra-PSU homogeneity and m is the number of sample PSUs.

4.3. Estimation

To estimate cost-efficiency as efficiency for set cost among cutoff options, we first solved for d_{c1} in Equation (6) to determine the number of responding dwellings for each option. Setting $\mathrm{Cost}_c = \mathrm{Cost}^*$, for $c = 1, 2, \ldots, 8, 15$, we determined the number of responding dwellings that could be afforded with Cost^* available under the cth option as,

$$d_{c1}^* = \frac{\{\mathrm{Cost}^* - F - T_c^*\}}{\{\mu_{c1} + \mu_{c0}(1 - \Gamma_c)/\Gamma_c\}}$$

where $T_c^* = T_c\{(\mathrm{Cost}^* - F)/(\mathrm{cost}_c - F)\}$ is an imputed value for T_c; Γ_c is the proportion of NHIS sample dwellings that are respondents as of the cth option; and μ_{c1} and μ_{c0} are weighted estimates obtained from cost study data on S_c and U_c, respectively. An estimate of Γ_c was obtained for the cth cutoff as the product of the actual NHIS dwelling response rate, computed as the number of respondents as of the cth attempt divided by the number of eligible dwellings in the sample, and a weighted estimate of the cutoff-specific eligibility rate for dwellings, defined for each cutoff from the cost study data as the ratio of the number of eligible dwellings to the number of sample dwellings. Because d^*_{c1} is a count of dwellings, we must convert it to a count of sample persons in a subpopulation (call it n^*_c) by multiplying d^*_{c1} times the average number of such persons per responding dwelling under the cth option, as obtained from NHIS respondent data.

With the person sample size for a cutoff at c attempts set at n^*_c when $\text{Cost}_c = \text{Cost}^*$, the expected mean squared error of ϕ_c for this group of respondents was then obtained by using direct estimates of $\text{Bias}(\phi_c)$ and anticipated values of $\text{Var}(\phi_c)$ using parameters estimated from the variance model. To estimate the increase in the variance of ϕ_c due to variable sampling weights (Deff_{Wc}) we used

$$\text{deff}_{Wc} = 1 + rv_{Wc} \tag{9}$$

(Kish 1965, sec. 11.7), where rv_{Wc} is the relvariance of sampling weights recomputed for the set of n_c respondents used to produce ϕ_c. Solving for Roh_{Ic} in Equations (7) and (8), we used deff_{Wc} from Equation (9) and deff_{oc} from estimates of variance obtained for ϕ_c from NHIS respondent data to obtain an estimate of Roh_{Ic} as

$$\text{roh}_{Ic} = (\text{deff}_{oc}/\text{deff}_{Wc} - 1)/(n_c/m - 1).$$

Because of demonstrable instability in subpopulation estimates, the value of roh_{Ic} from the corresponding total population estimate was imputed as the estimate of roh_{Ic} for subpopulation estimates of Φ. The set-cost variance was then computed as

$$\text{var}^*(\phi_c) = \text{deff}^*_{oc} s^2 / n^*_c$$

where

$$\text{deff}^*_{oc} = \text{deff}_{Wc}\{1 + \text{roh}_{Ic}(n^*_c/m - 1)\}$$

and

$$s^2 = n_{15}\text{var}(\phi_{15})/\text{deff}_{o,15}.$$

We estimated s^2 using NHIS respondent data from the current practice option, since n_c is greatest at this cutoff and thus s^2 will be the most stable among cutoff options. Finally, for a respondent sample of size n^*_c when total cost is set at Cost^*, the mean squared error of ϕ_c was obtained as

$$\text{mse}^*(\phi_c) = \text{var}^*(\phi_c) + \{\text{bias}^*(\phi_c)\}^2 \tag{10}$$

where $\text{bias}^*(\phi_c) = \text{bias}(\phi_c)$, the direct estimate of bias obtained from Equation (2), since bias is not affected by n^*_c. These findings are presented as averages of a relative-to-minimum squared root of the estimated mean squared error ($\text{rmse}^*(\phi_c)$; i.e., as

$$\text{rrmse}^*(\phi_c) = \text{rmse}^*(\phi_c)/\underset{c}{\text{Min}}\{\text{rmse}^*(\phi_c)\}. \tag{11}$$

5. Findings

5.1. Cost and efficiency

Two of the plots in Figure 1 describe how NHIS operational measures, the overall response rate among sample members and total cost relative to current practice, are affected by the choice of the number of allowable call attempts. As expected, both are nondecreasing with a rate of increase that diminishes as c increases. Moreover, note the similarity of the patterns among options for these two measures, thus indicating a strong linear relationship between cost and respondent sample size, because a plot of respondent sample sizes would follow the same pattern as the indicated plot of response rates.

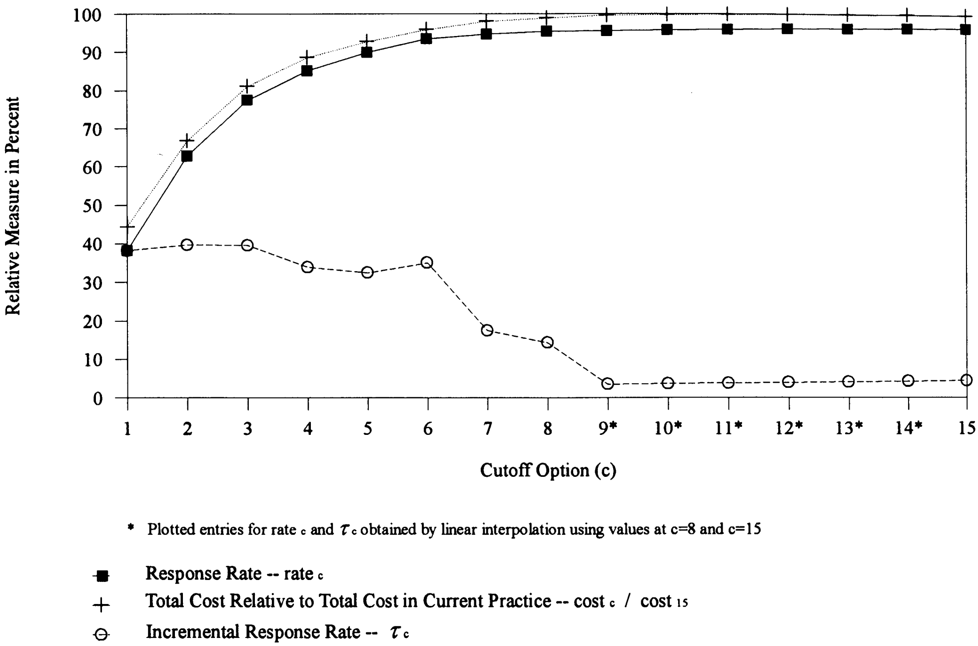

Fig. 1. Surveywide operational measures among cutoff options

The third plot in Figure 1 is for the incremental response rate, defined for a cutoff at c call attempts as $\tau_c = (n_c - n_{c-1})/(n_e - n_{c-1})$, where n_e is the number of eligible persons in the entire NHIS sample. Plotted values between $c = 8$ and $c = 15$ are obtained from a linear interpolation of the respondent sample size, n_c. Values of τ_c indicate the likelihood of obtaining a response from the remaining nonrespondents on a given call attempt and therefore reflect the interviewers' success in obtaining a response, given another call attempt. These findings reveal that this success rate drops only slightly through the first six attempts but then steadily drops further through the remaining attempts, thus indicating increased difficulty in obtaining a response with each additional attempt.

The relatively high level of interviewing success through the first several attempts is not surprising. NHIS interviewers are trained to gather pertinent nonresponse information (e.g., best time to contact) during each call attempt and to apply it to solicitation efforts on subsequent attempts. Adamant refusals, chronic not-at-homes, and other low-yield households are also likely to be eliminated on the first attempts, thus leaving the more willing but less accessible for later calls. Moreover, interview scheduling with less accessible households during the first call attempts can pave the way for continued success on later attempts.

Additional relevant measures of NHIS per unit cost are presented in Figure 2. There we see plots for the following by cutoff option: μ_{c1}, the average cost per responding dwelling for successful attempts; μ_{c0}, the average cost per nonresponding dwelling for unsuccessful attempts; and the incremental per respondent cost for successful attempts, $\delta_{c1} = (\text{cost}_c - \text{cost}_{c-1})/(d_{c1} - d_{(c-1),1})$, where $\text{cost}_0 = \$0$ and $d_{01} = 0$.

These results reveal several important cost patterns among cutoff options. First,

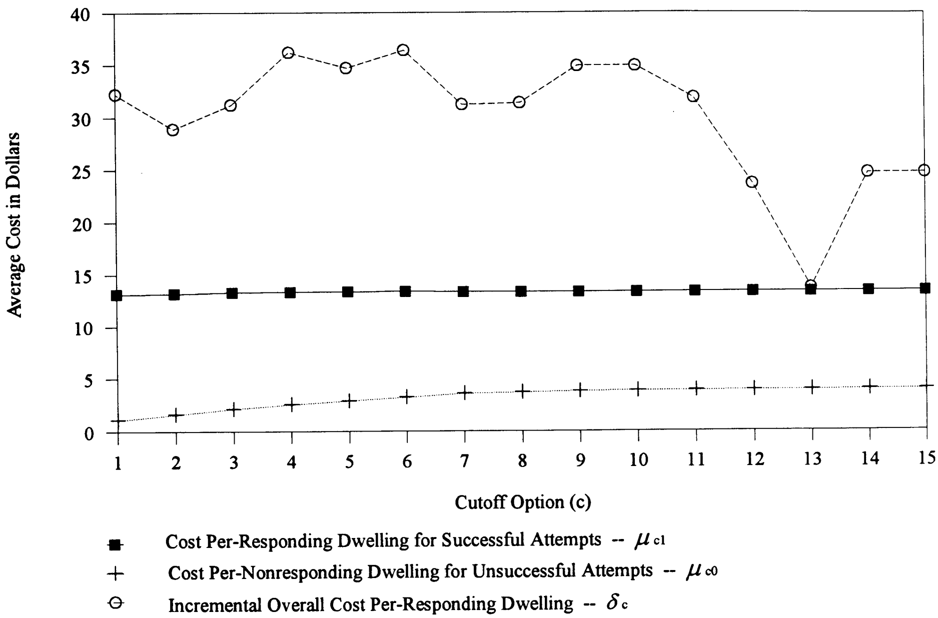

Fig. 2. Surveywide cost measures among cutoff options

the average interviewer salary per completed interview (μ_{c1}) varies only slightly among cutoff options and is substantially greater than the average interviewer salary per nonresponding dwelling for unsuccessful attempts. The explanation for these differing levels of unit costs is simple. Completing and editing an NHIS interview, with a core interview and several supplements on specific topics, require substantially more interviewer time than an NHIS nonresponse with only a few minutes required for each of a few call attempts. Second, although the average per nonrespondent cost is relatively low, it does increase dramatically from $c = 1$ to $c = 15$. This, of course, is not a surprising finding. Finally, although somewhat greater for $4 \leq c \leq 6$, there does not appear to be a strong relationship between c and the average incremental overall cost of a responding dwelling. Although at first glance this finding seems counterintuitive, given the trend for lower interviewer success (Figure 1) and higher per nonrespondent costs (Figure 2) as c increases. However, these two trends are not strong enough to counter the patterns of two large cost components, travel cost (T_c) and interviewing cost (S_c), which largely follow the pattern of respondent sample sizes.

The contents of Figure 3 profile changes in the demographic composition of the NHIS sample among cutoff options. For each demographic attribute the value plotted for a cutoff at c attempts is 100 times the unweighted percentage of the sample at that cutoff, divided by the comparable percentage under current practice. The findings reveal that sample composition does not change much with respect to the percentage of males, blacks, and persons in the 16 years and younger age groups combined. The sample composition, however, dramatically changes for other characteristics. For example, when $1 \leq c \leq 4$, and especially when $c = 1$, the respondent sample overrepresents the

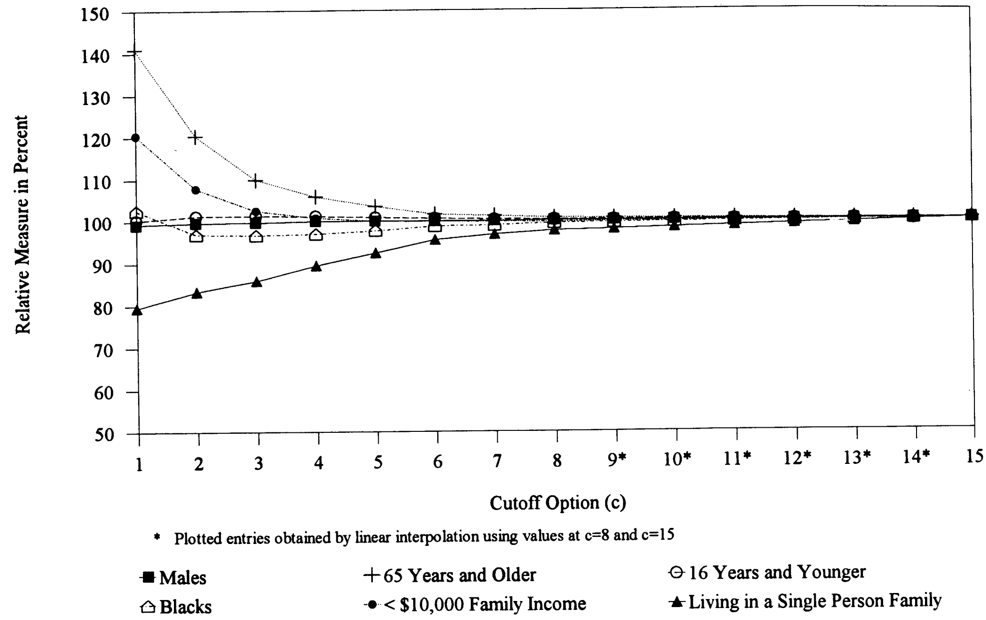

Fig. 3. Demographic profile relative to current practice of NHIS respondents by cutoff option

elderly and poor, and underrepresents persons from single-person families.

The findings in Figures 4 and 5 indicate typical patterns among cutoff options for the nonresponse bias, variance, and overall design effect associated with individual estimates of Φ. Results for the total population estimate of the proportion of persons reporting excellent health (Figure 4) demonstrate relatively stable trends for all three measures. Following the drop in the rate of nonresponse, the negative nonresponse bias of ϕ_c rapidly converges to zero over the first four options and then hovers closely about the line of convergence over the range of subsequent options.

Like the absolute value of the nonresponse bias, the variance of ϕ_c decreases over the range of options but at a more modest rate of decline. One important factor in this less dramatic drop is the steady increase in the overall design effect (indicated in Figure 4) that occurs as the average number of respondents per sample PSU increases with the higher response rate (see Equation 7). A somewhat sharper increase in deff_{oc} and milder decrease in variance would have been observed over c had it not been for the accompanying steady decline in the design effect due to variable weights (deff_{Wc}) from about 1.12 for $c = 1$ to 1.08 for $c = 15$.

The plots in Figure 5 are for NHIS estimates of the proportion of persons with a kindergarten education or less who are currently limited in their more minor activities. For this and most other subpopulation estimates we see the same general trends among cutoff options for the variance, overall design effect, and the absolute value of the nonresponse bias, except that the smaller sample sizes (n_c) for subpopulations tend to make the measures less stable and the patterns over c somewhat less apparent. For this

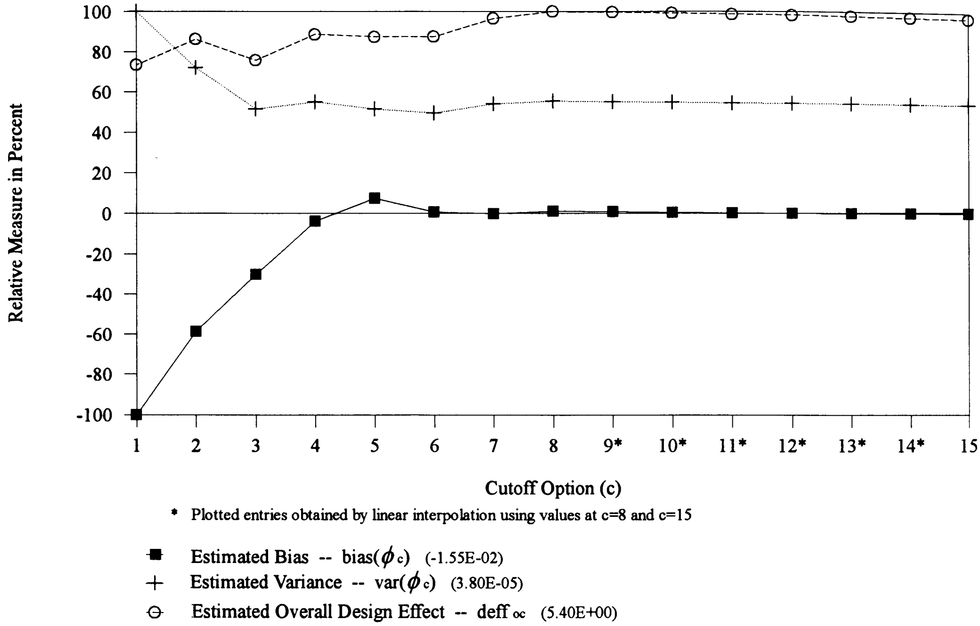

Fig. 4. Statistical measures relative to the absolute maximum among options for estimating the percent of the total U.S. population reporting excellent health status (maximum measures in parenthesis)

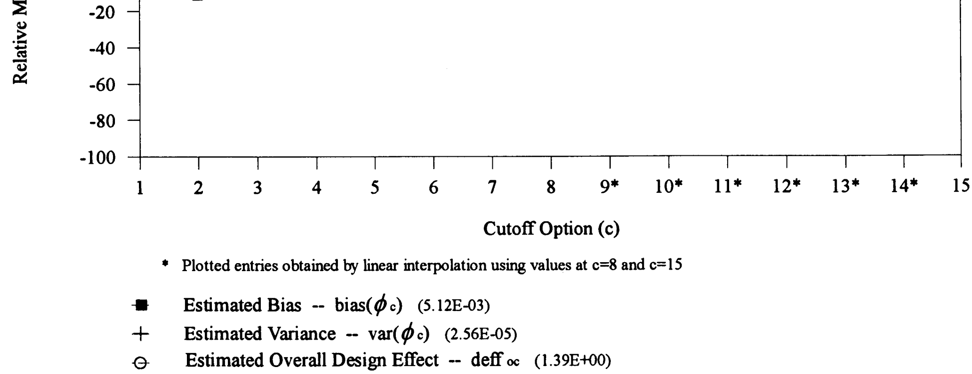

Fig. 5. Statistical measures relative to the absolute maximum among options for estimating the percent of the proportion of persons with a kindergarten education or less who are currently limited in their minor activities (maximum measures in parenthesis)

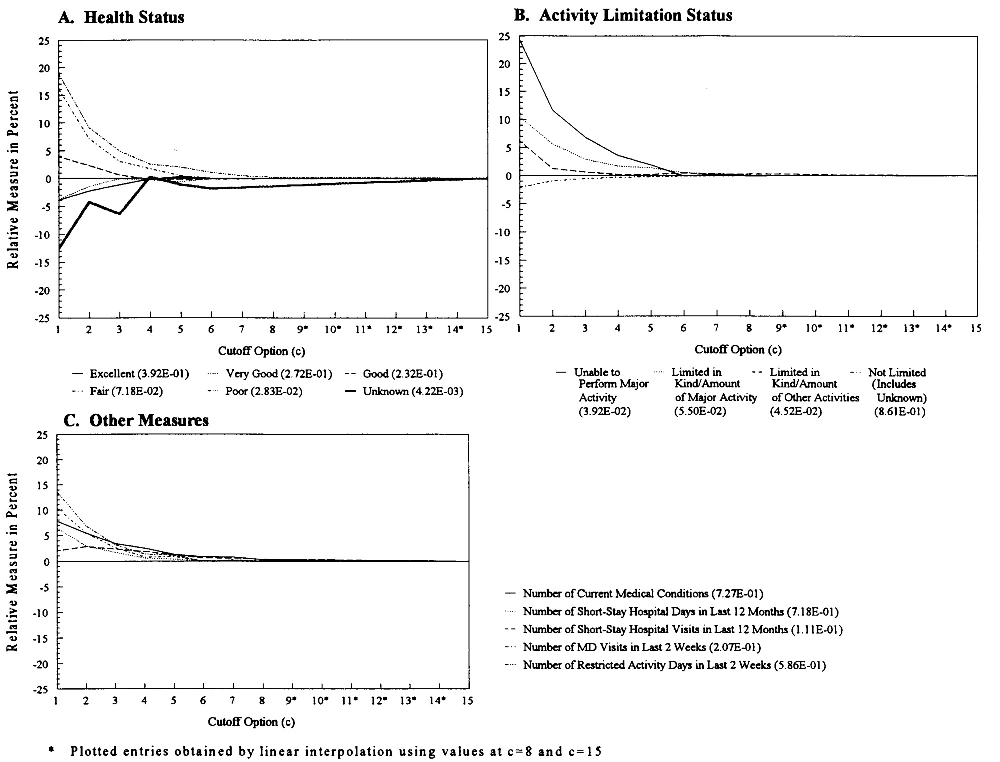

Fig. 6. Patterns of percent relative bias due to nonresponse by cutoff option for NHIS total population estimation (estimate from current practice sample given in parenthesis)

particular subpopulation estimate, the relative instability is especially evident for the nonresponse bias of ϕ_c, although there still exists some visual evidence of an inverse relationship between the magnitude of the bias and c. As with the total population estimates, the variance plot exhibits a downward trend as c increases with this subpopulation estimate, although the rate of decline is less due to smaller sample sizes $(n_{15} = 2,377)$.

Figure 6 displays trends in the percentage relative nonresponse bias of ϕ_c for the 15 NHIS total population estimates that were studied. Measures of rel-bias(ϕ_c), rather than bias(ϕ_c), are plotted to avoid scale issues. The rel-bias(ϕ_c) also measures the relative difference between the estimate of Φ obtained for a cutoff at c call attempts and the comparable estimate given the current practice of allowing up to 15 attempts (see Equation 3).

Although individual plots differ somewhat due to the varying sizes of Φ, all three sets of plots in Figure 6 indicate the predictable tendency for the nonresponse bias to diminish in absolute size as c increases. Because it is directly proportional to the rate of nonresponse (i.e., one minus the response rate), we expect, based on Figure 1, that the nonresponse bias in these plots converges rapidly toward zero for $c \leq 5$ but displays relatively little further reduction in size for $c \leq 6$.

The plots in Figures 6A and 6B are for estimates of population proportions associated with the sets of response categories to the NHIS questions on perceived health status and current activity limitation status, respectively. Since the proportions for each set of categories are functionally related (i.e., they sum to one), it is not surprising that some of the nonresponse biases in each set are negative while others are positive. For health status estimates tied to more positive self-perceptions (i.e., "excellent" and "very good") biases were negative, while those linked to the more negative assessments were positively biased. Altogether, these findings point to the general conclusion that the effect of nonresponse on estimates of health status produces a gloomier assessment of the health status of the U.S. population than what actually exists. The plots in Figure 6B indicate that nonresponse influences estimates of current limitation status by overstating the level of limitation in the population.

Finally, the plots for estimated means of various continuous measures in Figure 6C point to consistently positive biases due to nonresponse, although the magnitude of relative bias is generally somewhat less than for the two categorical items. The positive nonresponse bias indicated here might therefore be expected to at least partially cancel the typically negative measurement bias associated with health measures of this kind (NCHS 1977; Sikkel 1985).

The direction of the nonresponse bias seen in Figure 6 can be partially explained by the results in Figure 3. We see from Figure 6 that for $c \leq 5$ estimates of ϕ_c portray the population as: (i) seeing itself as less healthy, (ii) being more limited in its present activities, and (iii) being more likely to visit a hospital or doctor's office. These nonresponse biases are a direct result of the sample overrepresenting subpopulations that possess these same tendencies (i.e., the elderly and the poor) and underrepresenting those subpopulations whose likelihood of possessing these tendencies is relatively low (i.e., persons living alone) (see National Center for Health Statistics 1987).

5.2. *Cost-efficiency*

With Cost_c set for each cutoff option at $\text{Cost}^* = \text{cost}_{15}$, the estimated cost of the NHIS under the current option of allowing up to 15 call attempts, and using $m = 198$, the number of NHIS sample PSUs, estimates for a measure of the relative size of the mean squared error of ϕ_c are presented in Figure 7. The two plots in this figure, one for total population estimates of 15 reported NHIS health indicators and the other for 41 subpopulation estimates associated each of the same 15 indicators, profile the relative set-cost statistical quality of NHIS estimates among cutoff options. Each plot is for the average of the relative-to-minimum root mean squared error ($\text{rrmse}^*(\phi_c)$), with the smallest values for each plot indicating the greatest cost-efficiency among cutoff options.

It is apparent from the findings in Figure 7 that while the current practice is technically the most cost-efficient, any option among those with $c \geq 6$ will be almost as good. Starting at $c = 6$ the realized root mean squared error of ϕ_c will be within 6 percentage points of the lowest average $\text{rrmse}^*(\phi_c)$ that is obtained by allowing calling to continue through the 15th attempt. Findings in this figure also reveal similar patterns in average $\text{rrmse}^*(\phi_c)$ among call options for both total population and subpopulation estimates, although values of the former tended to be somewhat greater when $c \leq 4$. The lower average values of

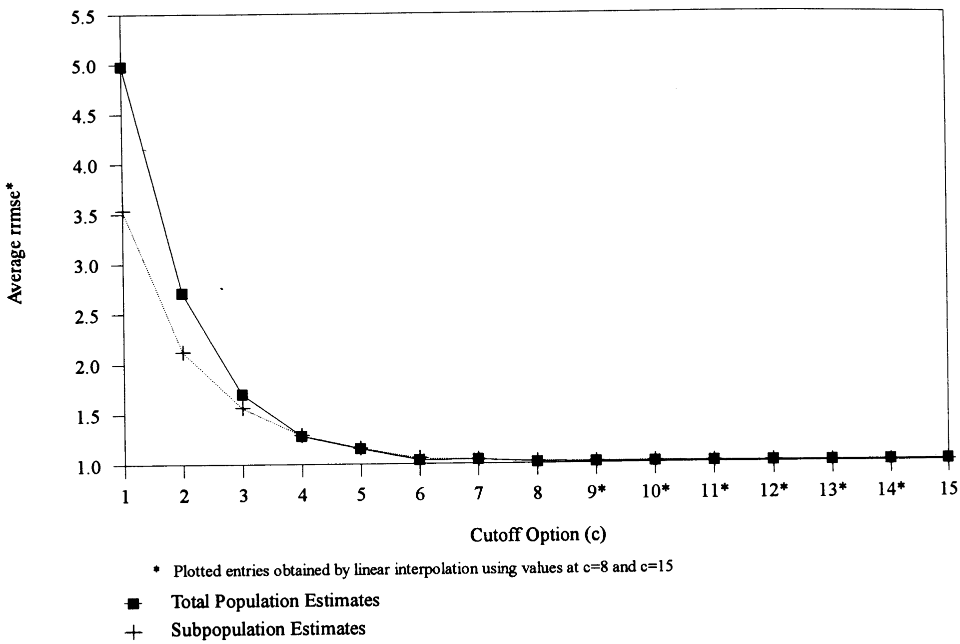

Fig. 7. Average root mean squared error relative to minimum by cutoff option, when cost is fixed at current level

rrmse(ϕ_c) for subpopulation estimates are due to the generally smaller nonresponse biases of these estimates relative to total population estimates and to the relatively minor role that variance plays in contributing to differences in mse(ϕ_c) among cutoffs. The nonlinear trend in rrmse$^*(\phi_c)$ for both plots reflects the strong influence of the bias of ϕ_s on the rrmse$^*(\phi_c)$ and the relatively minor role played by the variance of ϕ_c. The variance for set cost is less important here since with the cost set and per respondent costs of successful attempts (μ_{c1}) being virtually the same among options (Figure 2), the set-cost respondent sample size (n_c^*) and thus var$^*(\phi_c)$ are relatively constant among options.

Although not indicated in Figure 7, the current practice of allowing up to 15 call attempts was the option that most frequently produced the lowest mean squared error among options. The most cost-efficient value for c was 15 for 5 of 15 total population estimates and 446 of 615 subpopulation estimates (i.e., 72% of both types of estimates combined). Eight (8) call attempts was the next most frequent choice (13% of all estimates). The only subpopulations where 15 was not the consistent choice were the oldest, the least educated, and those living in large households.

6. Discussion

We have addressed the common and important issue of deciding how many call attempts should be allowed to obtain a response from members of a survey sample. The approach taken and illustrated by application to the National Health Interview Survey was to jointly consider the statistical and cost effects of competing options defined by the number of allowable call attempts. Comparing the cost-efficiency among cutoff options was done by setting costs for each option and then determining the expected statistical efficiency for each option. Relatively simple cost and mean squared error models are needed for this process. Findings revealed that while the current practice of allowing up to 15 call attempts is technically the most cost-efficient option among those considered, relatively little cost-efficiency is lost by reducing the allowable number of attempts to as few as 6 attempts.

The reported pattern of cost-efficiency among cutoff options (Figure 7) is generally due to the relative similarity of values of the overall per respondent cost of successful attempts (μ_{c1}) among options (Figure 1). Relatively large and roughly equal μ_{c1} translated into approximately equal set-cost sample sizes (n_c^*) and similar values among options for set-cost variances ($\text{var}^*(\phi_c)$). With variance virtually constant among cutoff options, the set-cost mean squared error was determined by the absolute size of the nonresponse bias ($\text{bias}(\phi_c)$), which in most cases was a decreasing function of c (Figures 4–6). This left the reported measure of cost-efficiency, $\text{rrmse}^*(\phi_c)$, to follow the pattern of the absolute bias, with rapid convergence through five or six attempts but only a modest decline thereafter until reaching a minimum at $c = 15$ (i.e., current practice). Thus one might speculate that a more dramatic rise in μ_{c1} over c might have led to a similar rise in $\text{var}^*(\phi_c)$ and, with the decline in $\text{bias}(\phi_c)$ over c, maximum cost-efficiency at an option other than the current practice.

6.1. Assessment of our approach

Data from the NHIS were clearly appropriate for assessing the cost-efficiency of cutoff options. Relatively good direct estimates of efficiency could be obtained from subsets of the full sample because of the high response rate for the NHIS under its current cutoff policy, although producing these estimates required reweighting the subset of the NHIS sample responding after c attempts. Useful estimates of cost were possible because of the detailed cost accounting of call attempts for a sizable NHIS subsample and because of our ability to isolate components of the cost of solicitation and data collection through the cth attempt. Although cost and efficiency estimates could not be derived for the same wave of the NHIS due to unavailability of 1987 NHIS data at the time of analysis, this limitation is not a major problem because of the relative uniformity of the NHIS survey design and related parameters over time.

Our empirical approach to cost-efficiency assessment was not without its limitations. One is the general presumption that the effects of a less stringent callback policy can be measured by simply truncating the number of call attempts under the current practice and observing interviewer efforts up to the truncation point. A reduction in the number of allowable call attempts might in reality also bring about changes in the response solicitation protocol that would affect efficiency in ways that this approach cannot fully detect. For example, reducing the number of allowable attempts to, say, three might force field staff to make more effective use of the three available opportunities than one observes by looking at the first three attempts of the current cutoff option of the NHIS, where interviewers know that many more solicitation attempts can be made after the first three.

A second possible restriction of our approach is that a portion of the NHIS operating costs that would be influenced by the size of the sample was not accounted for in the cost model. Our findings might have been altered somewhat if the costs of supervision as well as interviewer training and sampling activities had been included. However, as long as both the included and excluded costs are directly proportional to respondent sample size, any exclusion will not substantially influence our findings, provided the set level of cost ($Cost_*$) is understood to exclude these same costs.

Another limitation, related to the cost model, was our inability to isolate operational unit costs (μ_{c1} and μ_{c0}) for subpopulations. Our approach was to determine the expected total number of set-cost respondent dwellings under each cutoff and then to estimate the number of subpopulation respondents based on experience from the NHIS. Thus, findings for each subpopulation must be interpreted as those for a subpopulation within the context of a total population survey aimed at many subpopulations together, rather than findings that could be used to design a survey of the subpopulation by itself.

Applying the decision-making process of survey design to many subpopulation estimates leads to the inevitability of conflicting choices due to varying patterns of cost-efficiency among estimates, and a fourth potential limitation of our approach. Existing work on the problem of designing multipurpose surveys has focused more on how to combine the several optima from individual estimation settings and less on how to combine multiple estimation needs into a single optimization framework and solution (see, for example, Dalenius 1957; Dunkelberg and Day 1973; Kish 1988). One strategy suggested from this work is to choose as the global optimum that cutoff (c) with the smallest importance-weighted average of some measure of statistical efficiency among subpopulation estimates. We followed this approach but with equal importance weighting separately applied to each member of the sets of total population and subpopulation estimates. Thus, because the NHIS is a multi-purpose survey, no attempt was made to prioritize NHIS estimates; simple averages of efficiency measures were used.

Finally, because of the high NHIS response rate, we assumed that the remaining bias due to nonresponse for the estimate from the current cutoff option (ϕ_{15}) is negligible. If this assumption were not reasonable, the values of nonresponse bias(ϕ_c) would be off by a constant amount for each option. Since relative cost-efficiency among options was largely determined by the pattern of nonresponse bias, any bias in the bias measures would not have changed our conclusion that the current practice is the most cost-efficient cutoff option but it might have altered the patterns of convergence for our measure of cost-efficiency (rrmse$^*(\phi_c)$).

6.2. *Other variations of cost-efficiency measured by fixing cost*

Assessing cost-efficiency for design parameters in survey design is not a strictly defined process, but rather one where a diversity of plausible variations might be implemented around a common theme. The many procedural decisions we made as part of our set-cost approach to measuring cost-efficiency raises the question as to whether other variations arising from factors that entered into these decisions might have led us to findings that differ appreciably from those that were presented earlier. Because of this question, we found it helpful, and ultimately reassuring, to identify those factors that might have influenced our findings, to devise other approaches for measuring set-cost cost-efficiency, and to examine how, if at all, these other variations would have led to different conclusions from our original findings. The factors we considered are:

(i) the approach taken to estimate survey cost under each option and (ii) the level at which cost is set.

Our approach for estimating cutoff-specific cost (cost_c) was to project that portion of the total cost of the NHIS would have been needed if the cutoff for call attempts had been set at c. This was computed as the sum of the salary for all allowable attempts and the cost of interviewer travel for trips that would have been required under the option. A simpler approach in which segment- and assignment-level costs are uniformly apportioned to sample visits affected by each option was also considered and produced a similar pattern of estimated total cost and cost-efficiency among options to the approach we ultimately adopted.

Because the level at which costs are set, Cost^* influences n_c^*, deff_{oc}^*, and thus $\text{rrmse}^*(\phi_c)$. It was thought that the size of Cost^* might also affect the findings of our cost-efficiency analysis. To investigate this potential effect, we recalculated rrmse^* with Cost^* lowered to a level that would produce an 85% response rate. As indicated in Figure 1, the revised value of Cost^* was about 11% less than the value of Cost^* that was used in the original analyses. Our findings (not presented) demonstrate that this factor, like the other, has a minimal effect on the reported patterns of average rrmse^*. Values of average rrmse^* with a smaller Cost^* are consistently lower due to smaller n_c^* and deff_{oc}^*; however, only for $c = 1$ do differences in average rrmse^* for either total population estimates or subpopulation estimates exceed 0.15, thus making the patterns and implications of cost-efficiency for this lower level of set cost virtually identical to those from the higher set cost level.

Much of decision making in survey design depends on both statistical efficiency and survey cost, though the role they play is often quite different. Moreover, their roles may lead to countervailing influences, where an increase in one causes a reduction in the other. The challenge then is to recognize their roles in a design problem and to find ways to factor both into the decision. Concerning the matter of how many attempts are best allowed when soliciting participation in the National Health Interview Survey, we have seen that maximum cost-efficiency is not necessarily the only criterion one might use in choosing among options. The choice of which is best overall depends on what is being estimated, to whom the estimate applies, and what approach is followed in considering cost and efficiency together. Indeed, several options may be nearly as good as the best option. Clearly more thought is needed concerning ways to make these choices, but it is apparent that continuing the search for more operationally efficient methods of soliciting participation in surveys will make these decisions easier.

7. References

Cochran, W.G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley.

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, 164–169.

Dalenius, T. (1957). Sampling in Sweden. Stockholm: Almquist and Wicksell.

Drew, J.H. and Fuller, W.A. (1981). Nonresponse in Complex Multiphase Surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association, 623–628.

Drew, J.H. and Ray, G. (1984). Some Empirical Investigations of Nonresponse in Surveys with Callbacks. Proceedings of the Section on Survey Research Methods, American Statistical Association, 560–565.

Dunkelberg, W.C. and Day, B.S. (1973). Nonresponse Bias and Callbacks in Sample Surveys. Journal of Marketing Research, 10, 160–168.

Fitch, D.J. (1985). Sample Size Versus Call-Back Efforts: An Empirical Approach. Proceedings of the Section on Survey Research Methods, American Statistical Association, 522–524.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley.

Kish, L. (1965). Survey Sampling. New York: John Wiley.

Kish, L. (1988). Statistical Design for Research. New York: John Wiley.

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, Ser. 2, No. 110.

National Center for Health Statistics (NCHS) (1977). C.F. Cannell, K.H. Marquis, and A. Laurent: Vital and Health Statistics, Ser. 2, No. 69. DHHS Pub. No. (PHS) 77-1343. Public Health Service, U.S. Government Printing Office: Washington, D.C.

National Center for Health Statistics (NCHS) (1987). D.A. Dawson and P.F. Adams: Current Estimates from the National Health Interview Survey, United States, 1986, Vital and Health Statistics, Ser. 10, No. 164. DHHS Pub. No. (PHS) 87-1592. Public Health Service, U.S. Government Printing Office: Washington, D.C.

Rao, P.S.R.S. (1982). Callbacks, an Update. Proceedings of the Survey Research Methods Section, American Statistical Association, 85–87.

Rao, P.S.R.S. (1983). Callbacks, Follow-Ups, and Repeated Phone Calls. In Incomplete Data and Sample Surveys, Vol. 2, Theories and Bibliographies, edited by Madow, W.G. and Olkin, I., New York: Academic Press.

Shah, B.V. (1981). SESUDAAN: Standard Errors for Survey Data Analysis. RTI/5250/00-01S, Research Triangle Institute, Research Triangle Park, NC.

Sikkel, D. (1985). Models for Memory Effects. Journal of the American Statistical Association, 80, 835–841.

Steeh, C.G. (1981). Trends in Nonresponse. Public Opinion Quarterly, 45, 40–57.

U.S. Bureau of the Census (1985). National Health Interview Survey Interviewer's Manual. HIS-100(1985), Washington, D.C.

Zarkovich, S.S. (1963). Sampling Methods and Censuses; Vol. 2, Quality of Statistical Data. Rome: FAO.

Received June 1990
Revised March 1994