COST FUNCTIONS FOR NHIS AND IMPLICATIONS FOR SURVEY DESIGN

David Judkins and Joseph Waksberg, Westat, Inc.; Doris Northrup, CODA
1650 Research Blvd., Rockville, MD 20850

KEY WORDS: Sample Design, Optimization

1. Introduction

A classical problem in survey research is how to optimize sample design with respect to variance and cost. The variance properties of various sampling procedures have always been of great interest to mathematical statisticians, but concern about costs is another common thread running through the last 60 years of survey methods research. Despite this concern, it is rare for the major survey administrators to have adequate cost models for multi-stage area surveys of the complexity actually fielded. The cost models that are available for multi-stage area surveys are usually too simplistic. Shells for more complex models have been developed but have been repeatedly stymied by a lack of data to fit them. The recent book by Groves renews the call for the development of sophisticated models and points out a number of problems overlooked by the traditional models, for example, that survey costs are often nonlinear functions of sample size, that they are usually discontinuous over large sample ranges, that they vary stochastically, and that cost data have a very limited domain of applicability, that is, they are not generalizable from one survey to another. In the work reported on in this paper, we encountered not only the kinds of limitations described by Groves, but also many difficulties in using existing cost data from an ongoing survey to construct a model.

Survey field organizations keep track of costs for two primary purposes—paying field staff and monitoring field expenditures against the appropriate line items in the budget. Neither of these objectives necessarily requires that the cost of individual cases be known. The cost modeller, on the other hand, would like to associate costs with individual cases or at least with small sets of cases, say, segments.

It is understandable that survey organizations do not try to accumulate costs at the case level. First, it requires designing a considerably more complicated approach to recording time and expense, training interviewers to use it, and then monitoring interviewer performance in using it. Second, in hiring, training, and evaluating interviewers, survey organizations emphasize a number of essential qualities. Most important among these are the willingness and ability to go into unknown neighborhoods and knock on unfamiliar doors; to meet and persuade total strangers of the merits of their mission; to follow instructions carefully and read questions aloud carefully; to record answers correctly, and—last, but not least—to maintain rapport with respondents throughout this process. Not surprising, perhaps, many people who do well in these areas of primary importance to interviewer performance do not seem to do well at complicated bookkeeping tasks.

In this paper, we report on what turned out to be two fairly separate lines of research. The first line concerns our attempt to fit a fairly complete model for the costs of the National Health Interview Survey (NHIS) using administrative data and special cost studies. The second line concerns our use of simple cost considerations to approximate optimal design for NHIS. We had hoped to be able to use the formal model in the optimization process, but there were numerous difficulties. We think that both the difficulties experienced in fitting complex cost models and the optimization effort with simple cost considerations will be instructive to other researchers. More complete detail on these subjects and information on other aspects of the research on NHIS can be found in the complete report to be submitted by Westat to NCHS later this year.

In Section 2, we describe the sample design options which were under consideration and the basic interactions of cost with design. In Section 3, we discuss our attempt to fit a complex cost function. As will be seen, the available data did not appear to be adequate for the purpose of modeling costs. The results were not in accord with conventional survey wisdom or our own experience concerning the effect that various survey features have on cost. In Section 4, we recount the analyses leading to costs and operating efficiencies of alternate sample designs and how we proceed in the absence of a workable cost model. In Section 5, we close with some general guidelines for the optimization process and some suggestions on how to collect better cost data in the future.

2. NHIS Methods That Affect Costs

2.1 Sample Design Options

The NHIS is a three-stage probability sample of households. The first stage (PSUs) consists of individual counties and groups of contiguous counties. Within a PSU, two types of second stage units are used: area segments and permit area segments. Area segments consist of blocks, groups of blocks, or enumeration districts. Permit segments are used to represent units built after the census prior to the most recent redesign. They are based on lists of building permits issued after the date of the census. The third stage consists of households within the second stage units. Data are collected about all members of sample households.

We will describe the results of research into the cost and sample design implications of different ways of improving the reliability of statistics for demographic domains. These domains were specified by age (6 categories), sex, race (black versus other), and ethnic origin (Hispanic versus other). The methods considered involved one or more of the following features: increase the number of PSUs, increase the total sample size, introduce screening, introduce subsampling within households, introduce the use of special frames, oversample areas with strong minority concentrations, and introduce network sampling. The sample size for screening might be obtained either by increasing the designated cluster size or by increasing the number of clusters.

Our analyses are restricted to data collection costs, that is, those incurred by the interviewers or listers. Although more complex sample designs will require more effort in sample selection, sampling costs appear to be such a small part of the total that they can be ignored in decisions on sample design.

2.2 Interview Methods That Affect Costs

Survey costs are affected by the operational plan of the survey as well as the volume of work. In the case of NHIS, there are four major operational factors influencing the costs:

The deployment of the field staff. The current NHIS operational plan utilizes a field staff of resident interviewers in all PSUs with large workloads and in roughly half of the other PSUs. Nonresident interviewers are used for the balance of PSUs with small workloads. When possible, nonresident interviewers are strategically positioned so that PSUs without a resident interviewer can be reached as efficiently as possible, usually by use of personal automobile without incurring the cost of air travel. If interviewers run into difficulty finishing their work within the allotted time in "away PSUs," there is a cadre of more experienced and higher paid personnel, referred to as Supervisory Field Representatives (SFRs), who often can be brought in to help finish assignments in addition to their normal quality control or other supervisory activities.

The definition of an assignment. The sample in a PSU is divided into a set of assignments, each of which was originally expected to provide an adequate one-week workload for an interviewer. The assignments are randomly allocated to weeks. The plan established in the latest NHIS redesign was to give each interviewer about six to eight assignments per quarter with about 16 households per assignment. Because of the increased length of supplements the average assignment now takes 51 hours to complete. As a result of the increasing interview lengths, many assignments cannot be completed in a single week. The ideal is to give an assignment to an interviewer only every other week. This keeps them on a part-time basis and allows them a cushion if they have trouble finishing the assignment in the assigned week. However, this has not always been possible because of the random distribution of assignments across a quarter. Although it is preferred that the assignment be finished within two weeks, there is evidence from survey records that many cases are finished in the third week or beyond.

Interview factors. The NHIS core interview has remained essentially unchanged for many years. Although some changes are expected, once they have been made, the core questionnaire is expected to be fairly stable over a fairly long period of time. However, there are many supplements. These supplements change annually and are of varying lengths. Total administration time for the supplements has increased considerably over the past several years to the point where they now account for a substantial part of the total interview time.

The survey attempts to obtain data for the entire civilian noninstitutional population. All adult members of the sample households 17 years of age and over who are at home at the time of the interview are invited to participate in the interview and to respond for themselves. For children and for adults not at home during the interview, information is provided by a responsible adult 19 years of age or over, residing in the household. With respect to the supplements, some can be answered by the respondent, while others require self-response by a randomly selected sample person.

The implementation of the sample design. The Census Bureau has developed a procedure for the replacement of the sample from year to year that reduces the number of area segments that have to be mapped and listed each year to about 18 percent of the total number of area segments. The Census also has the option to maximize the overlap of NHIS segments with those of other current surveys to reduce further mapping and listing cost, but instead the overlap has been minimized to avoid possible confidentiality problems.

3. Cost Function and Estimates of Parameters

We initially tried to construct a cost model that would take the number of PSUs, the total interviewed sample size (targeted), the sample size to be screened, the method for obtaining the extra screener households, the proportion of persons within sample households to be interviewed, the proportion of the sample to come from special frames, the proportion of the sample to come from areas with strong minority concentrations, and the proportion of the sample to come from network sampling all as input and then output a cost estimate. However, we scaled our goal back as we evaluated the data available for model fitting.

3.1 Cost Function

We split the costs into four major components: PSU, segment, housing unit (HU), and within household (HHD).

PSU Cost. Most of the PSU component is caused by travelling interviewers. There is no contribution to this component by PSUs in which the work is performed by a resident interviewer. Currently, the NHIS field staff is comprised of about 60 supervisory field representatives (SFRs), all of whom work part-time on NHIS and part-time on other Census studies, and about 125 field representatives (FRs), some of whom also work on other Census studies. About 50 of the FRs work only in their home PSUs, while about 75 of them cross PSU lines (or State lines within Metropolitan Statistical Areas (MSAs)) to perform their work. These 75 travel to one or two other PSUs, generally by personal automobile, for which they are reimbursed for mileage and other expenses. Under extraordinary circumstances, the interviewer is permitted to incur air travel and the cost of a rental car. If the interviewer is out for 10 hours or more and is 50 miles or more from home, per diem is granted.

The cost of between-PSU travel is basically a function of how many PSUs must be serviced by travelling interviewers, how far the interviewers have to travel, and how long they have to stay. Factors to consider when predicting the number of PSUs which will require work by out of town interviewers include the physical size of the PSU (more hours spent travelling within the PSU make it more feasible to have a resident interviewer), the sample size, the questionnaire length, and the degree of clustering within the PSU. The average distance travelled depends on the number of PSUs in the sample. Distance roughly decreases with the square root of the number of PSUs. Regarding length of stay, the current NHIS practice is to stay one week at a PSU.

Segment Cost. The segment cost is affected by the need to map and list segments prior to final sample selection and by travel between segments. Also, within-segment travel can be substantial for permit segments and some rural area segments. Although it seems natural to view all travel between segments as segment cost, this travel is strongly dependent upon the number of households in each segment and on respondent rules. This is due to the fact that 5 or 6 visits to a segment are typically required to complete it. A requirement for self-response would obviously increase the likelihood of missing at least one sample person on the first visit. With empirical data on the number of trips per segment, it might be possible to model this relationship with hazards analysis. Since we, however, have no such data, we left the definition of segment cost deliberately vague, saying simply that it is the part of cost that responds linearly to the number of segments. According our model is simply

$$\text{segment cost} = C_{seg} \text{ (\# segments)}.$$

If list samples are used as part of a dual frame approach, then there would essentially be two very different kinds of segments. Each sample unit from a list could be regarded as a separate segment. The list segments would then not require listing and would need fewer trips per segment to complete. It would thus be necessary to have two segment cost components, one for each type of segment.

Housing Unit Cost. Costs associated with HUs are those incurred after arriving at the segment including part of the cost of repeat visits to the segment. Thus the cost associated with a HU is the cost of knocking on a door, determining eligibility of the structure (including administering any screener questionnaire), selecting a within-HHD sample (if applicable), interviewing one person, and part of the cost of multiple visits.

The HU cost can thus be modeled as

$$\text{HU cost} = C_{HU1} \text{ (\# HUs screened out)} + C_{HU2} \text{ (\# HUs with 1+ sample persons)},$$

where C_{HU1} is the cost of locating and screening out one HU once the interviewer has arrived in the segment, and C_{HU2} is the cost of locating and screening an eligible HU and of interviewing the first sample person. The values for the parameters C_{HU1} and C_{HU2} will obviously depend on the complexity of the screening and the length of the interview (including supplements). Since screening is not currently used, this may be simplified to

$$\text{HU cost} = C_{HU} \text{ (\# interviewed HUs)}.$$

Within-Household Cost. The within-HHD cost is the cost of obtaining the required information for the remaining sample persons within a HHD with at least one interview. For the current design, this cost is quite small even with several supplements. Note, however, that proxy response is currently allowed for the core questionnaire and for many of the supplements, but some of the supplements require self-response. Any changes to more strongly encourage interviewers to obtain self-response would affect both between-segment and household costs by requiring more trips to the segment. We mention this here since others might reasonably consider the travel costs of repeated visits to the same household as purely within-HHD cost. Our model for the within-HHD cost is

$$\text{Excess person cost} = C_{WHHD} \text{ (\# interviewed persons} - \text{\# interviewed HHDs)},$$

where C_{WHHD} is the cost of obtaining information about an additional person in a multi-person HHD.

3.2 Estimation of Parameters from Administrative Records

There are basically three approaches for obtaining cost data to fit models such as ours. The first is to conduct a special experiment using multiple methods with separate accounting systems for the different experimental treatments. This is very expensive and time consuming. The second is to institute the use of much more detailed cost report forms than are normally employed. This approach was tried by the Census in 1987. We reviewed these data but found that the objectives of the study had been too limited to support our needs. The third is to sift through the existing cost data that are kept for administrative purposes. We pursued the third approach.

Let us first review the sorts of records kept by Census. The Census records system was created for a different purpose than what we needed for cost modeling. The Bureau keeps data on payroll charges, mileage charges, per diem, telephone charges, and all other expenses, all detailed type of operation such as interviewing, listing, observation, reinterview, training, and miscellaneous other categories. They do not keep data on regular basis on the number of interviews that are done by supervisory field representatives as opposed to regular interviewers. They also do not keep data on the number done by resident interviewers as opposed to travelling interviewers. They do not keep records on the length of time to close out an assignment. They do not keep track of the mileage expenses incurred by field staff travelling between PSUs versus the mileage expenses incurred within PSUs. They do not keep track of interviewer time spent travelling as opposed to actual interviewing, since for payroll purposes it is not necessary to do so.

The Census Bureau provided us with a file of all payroll transactions related to NHIS. Air fare and rental car costs were not used in any of the analysis since we were informed that these costs primarily pertain to training, observation, and reinterview. NCHS provided us with an NHIS data file for the same period that contained all interview data, including some data items not normally released to the public, such as interviewer ID, length of interview in minutes, and number of calls required to complete the interview. By linking these together, we hoped to overcome the nonspecificity of the payroll charges. The challenge was relating charges to product. In this section, we describe some of the interesting problems that we encountered in trying to link costs with product. The effort was only partially successful. Thus, although there are some implications for NHIS redesign, the stronger implications probably concern record-keeping procedures.

The conceptual link between cost and product is the interviewer assignment. On the cost side, the interviewer is requested to submit charges every two weeks. On the other side, as each questionnaire is completed, the interviewer is supposed to enter his/her identifying code and the date that the interview was completed. Assignments are generally given every other week.

Interviewers are encouraged to complete the work within the assigned week or as soon thereafter as possible. Thus, by limiting our analysis of costs to those incurred during the first week of each payroll period, we thought it would be feasible to relate the characteristics of the interviewer's assignment (number of segments, households, persons) with the charges submitted for the payroll period. Multivariate regression techniques could then be employed to extract the marginal costs associated with segments, HUs, and persons.

The most obvious weakness with this approach is that a high percentage of assignments are not completed within a week and a substantial number are not completed within two weeks; this leads to interviewers frequently working on overlapping assignments.

Another problem is that there are staff that work on the survey without actually completing interviews. These include mostly senior field representatives (SFRs) that do quality control work. Recognizing these weaknesses, some noise was expected. Beyond this obvious weakness, various types of errors also introduced noise.

One problem that emerges in an analysis of survey cost occurs when survey projects share field resources. Occasionally an interviewer working on more than one project accidentally charges time or expenses to the wrong project. In looking at the relationship between cost and product, however, such errors introduce additional noise. There is also the problem of the late time and expense sheet. This is a fairly common problem in all survey organizations. In an analysis in which time is used as the basis of linking the cost data to the production data, however, it adds to the problems.

Another category for the analysis was coding and keying errors. Since the files were being linked on payroll period, regional office, and interviewer, any imperfections in these codes leads to false or ambiguous matches. Problems of this type occurred fairly often, particularly problems with interviewer IDs. We edited as best we could and dropped whole assignments when there was any doubt that we would be obtaining an exact match.

A final problem was posed by interpenetrated segments, i.e., segments worked by multiple interviewers. There are a number of reasons why survey organizations send more than one interviewer to a given segment—special language or other respondent needs, interviewer illness or attrition. Unfortunately, for the cost modeller's purposes, the survey administrative data usually provided for only field worker ID, that is, one person is credited with the completion or charged with the nonresponse for each case. This further complicates the attempt to associate costs with product. In order to avoid this problem, we dropped all segments worked by multiple interviewers and we also dropped the rest of these interviewers' assignments for the same payroll period.

The payroll file indicates 5,831 assignments for the last half of 1988 and the first half of 1989. The survey file, on the other hand, indicated something like 3,500 assignments. Of these, we ended up with matched data on only 1,482 assignments.

3.3 Best Parameter Estimates

As a starting point, we took some basic counts from the payroll file. These counts were not adjusted for any of the peculiarities indicated in Section 3.1. They are shown in Table 1. As mentioned previously, we focused only on the listing and interviewing costs, believing the other costs to be fairly insensitive to the sorts of design changes being considered.

We also established that there were 8,189 segments in 1988, 62,106 assigned addresses, 50,061 eligible households, 47,485 interviewed households, 122,310 persons; that the average noncertainty PSU comes into sample 11.2 times per year; and that the average assignment was 2.3 segments or 13.5 households. The average yield (percent of designated addresses that yield interviews) was 94.9 percent.

We had originally planned to model labor charges, mileage, and per diem. However, after obtaining somewhat disappointing results for labor charges, we dropped plans for modeling mileage and per diem.

PSU cost. In attempting to break the total payroll costs into between- and within-PSU components, we used information provided to us by the Bureau of Census about the operational procedures as well as the cost data. The Census Bureau uses a field staff of 182 field representatives and supervisory field representatives to cover the 200 PSUs involved in the NHIS. One hundred and thirty-three of them have regularly scheduled assignments in more than one PSU; 78 of these have regularly scheduled assignments that require overnight travel; and the other 55 have assignments in other PSUs within a 50-mile round trip from home and do not require overnight stays. Almost all between-PSU travel is by car, but some of it—particularly the travel incurred by the supervisory field representatives—occurs by air travel.

Table 1. Annual costs for NHIS interviewers and SFRs (before overheads and excluding air travel and car rental) by operation and type of expense

Type of Expense **Operation**	**Labor**	**Mileage**	**Per diem**	**Telephone**	**Other**	**Total**
Listing	$ 89,000	$ 40,000	$ 3,000	$ 1,000	$ 1,000	$ 134,000
Interviewing	1,449,000	400,000	241,000	58,000	31,000	2,179,000
Keying/Office	22,000					22,000
Training	139,000	18,000	69,000	1,000	9,000	237,000
Reinterview and observation	62,000	14,000	14,000	2,000	2,000	95,000
All other (including pretests)	38,000	2,000	12,000		6,000	58,000
Total	$1,799,000	$474,000	$339,000	$62,000	$49,000	$2,725,000

We made assumptions that the average round-trip distance between PSUs was 246 miles for those cases requiring overnight stays, and 45 miles for those cases which do not. We assumed that field staff travel these distances at an average speed of 50 miles per hour and that each makes about 11 visits to other PSUs per year. This set of assumptions allowed us to derive an estimate of the number of hours spent travelling by car between PSUs by NHIS field staff each year.

No estimate of airfare for between-PSU travel was available from the Census Bureau. We were told by Census staff that air travel for this purpose was not a significant line item in their NHIS budget. We had no way to estimate the amount of time spent in air travel, so we simply treated all between-PSU travel as though it occurred by car. Using these assumptions, we estimated that of the annual costs of interviewing labor ($1,449,000 in the payroll data available to us) about 6% (or $87,956) was for between-PSU travel.

Segment, HU, and Person Cost. It was in this area that we expected the matched HIS/payroll file to be useful. The workload sizes are summarized below:

No. per assignment	Average	Range	Standard deviation
Segments	2.3	1 to 7	1.2
Households	13.5	1 to 44	7.2
Persons	34.6	1 to 130	19.9
Hours	51	1 to 194	28

There seemed to be enough variation to have a good chance of predicting hours charged for interviewing including travel but excluding listing, training, observation, reinterview, etc. Unfortunately, it turned out that the HIS data are not very useful in predicting payroll charges. We go through some of the models that we tried and then present a model that blends the results of the formal fitting with general experience in other surveys.

A simple regression model with parameters for an intercept, segments per assignment, households per assignment and persons per assignment provided a terrible fit. The R^2 was under 7 percent, and the intercept was huge. (The intercept implied that an assignment would take 36 hours even if there were no segments, households, or persons in it!)

Hypothesizing that the lack of fit was at least partially due to carryover of work from one pay period to another, we fitted another model with an intercept and parameters for average numbers of segments, households, and persons over the current and preceding pay periods. Fit was improved but still not acceptable. The R^2 was about 13 percent, the intercept was still too large (25 hours for an empty assignment), and the parameter for households was negative with a large standard error.

We tried adding covariates. We tried this first without averaging with the lagged assignment. The covariates were type of urban/rural/metro/nonmetro setting, average minutes per interview for the assignment as recorded on the front of the questionnaire (with some editing of extraordinary values), average number of calls for the initial interview or final nonresponse, and average number of additional contacts. The R^2 was 0.076, the intercept was still very large (23.4 hours), and each minute per household as recorded on the questionnaire led only to 16 seconds of charges per household.

We tried adding a random effect for the interviewer. This improved the R^2 to 0.34 but led to nonsensical values for the fixed covariates (some were negative with huge standard errors).

Finally, we tried covariates that were averaged with the lagged assignments. The R^2 was still low (0.152) and the parameter for interview length still seems too small, but the intercept looks reasonable. The model is shown in Table 2.

The problem comes in deciding how to incorporate the covariates into our basic model. There are terms like the intercept and the coefficient for central city of small SMAs which do not fit easily into the cost function.

$$\text{Cost} = \text{Between-PSU costs} + C_{seg}\ (\#\ \text{segments}) + C_{HU}\ (\#\ \text{HUs}) + C_{WHHD}\ (\#\ \text{excess persons}).$$

Folding the type of area parameter back into the intercept gives a fixed cost of 1.2 hours per assignment ($2.4 - 3.3 \times .367$). The calls and contacts account for $13.9 = (4.3)(7.8) + (3.8)(0.49)$ hours per assignment. We somewhat arbitrarily allocated half of that to segment cost and half to HU cost. Thus hours per segment is equal to $6.5 + (1/2)(13.9)/2.3 = 9.5$ and hours per household is equal to $0.8 + (1/2)(13.9)/13.5 = 1.3$.

Table 2. Regression model for hours charged by assignment

	Beta	SE	Avg.
Intercept (hours per assignment)	2.4	4.5	1
Central city of small MSA or urban suburb of large MSA	-3.3	1.4	.367
Average number of segments over last two assignments	6.5	1.5	2.3
Average number of interviewed HHs over last two assignments	0.8	0.2	13.5
Average minutes per interview over last two assignments	0.12	0.04	72.8
Average number of calls for initial interview or final nonresponse over last two assignments	4.3	0.9	2.8
Average number of additional contacts over last two assignments	3.8	1.9	0.49

We assumed that the minutes per interview were mostly due to excess persons per household even though a separate regression showed that persons per household explained less than 9% of the variation in recorded interview time (when both are aggregated to the assignment level). The hours per excess person is thus $\frac{(.12)(72.8)}{34.6 - 13.5} = .4$.

Assuming a flat rate of $8.28 per hour (includes overtime and night differential), the fitted model for the cost of interviewing becomes

cost = (1.2) $8.28 (# assignments)
+ $8.28 (9.5) (# segments)
+ $8.28(1.3) (# HUs)
+ $8.28(.4) (# excess persons)

= $9.94 (# assignments)
+ $78.66 (# segments)
+ $10.76 (# HUs)
+ $3.31 (# excess persons).

If we further assume that all $89,000 of labor charges (Table 1) for listing are segment costs, the labor cost of listing and interviewing becomes

cost = $9.94 (# assignments)
+ $89.53 (# segments)
+ $10.76 (# HUs)
+ $3.31 (# excess persons).

To validate this model, we substitute in the numbers of assignments, segments, HUs, and excess persons. This gives total labor cost for listing and interviewing of $1,527,000.

This agrees well with the total labor cost for listing and interviewing in Census' payroll records. Ironically now however, the fixed cost per assignment is lower than the number derived in the preceding subsection on between-PSU cost. There we estimated $87,000 compared to $35,000 here. Since the intercept in the model was subject to high sampling error, this is not too surprising. It is clear that the model is not very reliable. Relationships between product and charges have been too severely blurred by timing and recording issues to still be able to measure them accurately.

4. Interaction of Cost, Design, and Objectives of Survey

As we described, the cost models turned out to be quite rough approximations of how costs would vary with changes in elements of the survey design. We doubt that they are useful in choosing among minor variations in the sample design. However, an analysis of the implications of the most critical design decisions indicated that their effects were fairly clear even in the absence of a formal cost model. We describe the analysis of some of the principal sample design issues. They illustrate the interaction among costs, sample design, and principal survey goals, with recommendations reached in an iterative procedure. We believe this is a useful approach in many multipurpose surveys.

4.1 Effect of Screening and Interview Costs on Sample Design

As mentioned earlier, the major purpose of the research carried out was to determine the implication on costs and on the sample design of revising the sample to achieve a specified minimum precision for certain subdomains. There were several major types of subdomains. One comprised sex-age breakdowns for race/ethnicity groups consisting of blacks, Hispanics, and all others. This paper concentrates on research on these subdomains. We note that our estimates of population and sampling size relate to the year 2000, the midpoint of the time period for the redesigned NHIS.

We start with the sample sizes resulting from a self-weighting sample of 50,000 interviewed households, approximately the current sample size. The assumed population distribution for the year 2000 is largely based on Census Bureau projections, with some Westat estimates used to fill in missing data. We have also assumed that survey undercoverage by age-sex-race/ethnicity will follow the current pattern experienced by the Census Bureau in the Current Population Survey. Estimates of these sample sizes for the subdomains under consideration are shown in Table 3.

Table 3. Sample size by sex, age and race/ethnicity for a self-weighting sample of 50,000 interviewed households in the year 2000

	Total	Black	Hispanic	Nonblack, NonHispanic
Total	124,000	15,579	11,894	96,527
Males	60,056	7,198	5,921	46,937
Under 5	4,213	689	704	2,820
5-17	12,174	1,963	1,691	8,520
18-24	5,467	728	641	4,098
25-44	17,905	1,971	1,790	14,144
45-64	13,689	1,264	804	11,621
65 and over	6,608	584	290	5,735
Females	63,944	8,381	5,973	49,590
Under 5	4,007	660	674	2,673
5-17	11,568	1,873	1,615	8,080
18-24	5,606	852	620	4,135
25-44	18,594	2,500	1,747	14,346
45-64	14,609	1,632	911	12,065
65 and over	9,560	864	406	8,290

The sample sizes per subdomain vary from a high of over 14,000 for Whites, aged 25-44 to a low of 290 for Hispanic males 65 years and over. Although a self-weighting sample is efficient for statistics on the total population, the precision of data for subdomains will be extremely variable.

Consideration of the kinds of analyses that appear desirable for the subdomains and the design effects in the current NHIS indicates that the minimum sample size per subdomain necessary to achieve the desired precision is 1,000. As can be seen in Table 3, all of the nonblack, nonHispanic subdomains exceed 1,000, most by a considerable factor. Half of the black subdomains and one-third of the Hispanic sex-age subdomains meet this requirement. The smallest sample size is for Hispanic males, 65 years and over, where the sample is only 29 percent of the size deemed necessary.

Table 4 shows the percentage increase that is necessary for each subdomain with less than 1,000 sample cases, and the reduction possible in subdomains with over 1,000. The greatest increase, 245 percent is, of course, for Hispanic males, 65 and over. The next greatest one is for Hispanic females, 65 and over, which calls for a 146 percent increase. The greatest increase in blacks is for males 65 and over which requires a 71 percent increase.

Increasing the total sample by 245 percent just to achieve required subsample sizes for one or two of the subdomains is a very inefficient way to proceed. Consequently, the only practical way of meeting the survey goals is to start with a much larger sample, screen the households for sex, age, and race/ethnicity of the residents, and subsample accordingly. A new element of cost, screening, is thus introduced and this had to be reflected in the cost function.

Once screening is accepted as a necessary part of the survey procedures, it is sensible to examine sample designs that reduce the amount of screening. One effective procedure for accomplishing this is to sample areas with high concentrations of blacks and Hispanics at higher rates than the rest of the U.S. How to do this in a reasonably optimum manner is described in the literature, for example, by Kalton and Anderson,[1] and by Waksberg.[2] We note that this complicates both the cost and the variance function. The cost function is affected because, although the total screening is reduced, there is a shift in the geographic distribution with a disproportionate part in areas with high concentrations of minorities. Unit costs of screening in such areas may be different from those in the rest of the U.S. The variances are affected because oversampling in selected parts of the U.S. creates variability in sampling rates which changes the form of the expression for the variance.

The gains from differential sampling in high-minority areas are so great, however, that these complications are worthwhile. If a self-weighting sample for each subdomain is desired, it is necessary to screen almost 175,000 households to locate 1,000 Hispanic males, 65 years and over. We note that the Census Bureau estimates that the cost of screening a household which will not be interviewed is about one-third the cost of interviewing a household. The extra screening increases the data collection costs by more than 60 percent.

Table 4. Sample expansion or contraction for sex-age-race/ethnicity subdomains with a 50,000 household, self-weighting sample

Race, ethnicity and age	Males		Females	
	Reduction possible (%)	Increase needed (%)	Reduction possible (%)	Increase needed (%)
Black				
Under 5		45		52
5-17	49		47	
18-24		37		17
25-44	49		60	
45-64	20		39	
65+		71		16
Hispanic				
Under 5		42		48
5-17	41		38	
18-24		56		61
25-44	44		43	
45-64		24		10
65+		245		146
Nonblack, NonHispanic				
Under 5	65		63	
5-17	88		88	
18-24	76		76	
25-44	93		93	
45-64	91		92	
65+	83		88	

With oversampling in high minority areas, the screening is reduced to 120,000. This is accomplished by a moderate oversampling in most of the U.S. and high rates of oversampling in areas with high concentrations of blacks and Hispanics. Oversampling rates in areas with concentrations of Hispanics, which would achieve the desired precision or better in an efficient manner for all Hispanic subgroups, are shown in Table 5. A similar table would be used for blacks, with the maximum oversampling rates lower than Hispanics because smaller increases are needed. Very high levels of oversampling are necessary only for persons 65 and over. Other subdomains are oversampled at lower rates. The method of oversampling other Hispanic subdomains is such as to subsample the persons identified in the screening operation in a way that creates as little variation in sampling rates as is possible.

4.2 Alternatives for Rare Subdomains

If we look at the individual subdomains in Table 5, it can be seen that one subdomain, Hispanic males 65 and over, accounts for a significant part of the need for the large screening workload. Hispanic females also need considerable oversampling. If blacks and Hispanics 65 years and over could be handled in a different way, the screening would be reduced from 120,000 to 65,000. The cost would then be reduced by about 20 percent. We have accordingly explored other methods of dealing with these subdomains. At present, several alternatives are being considered. No decisions have yet been made since operational and other potential problems are still being examined.

1. One alternative is dual-frame sampling. The HCFA medicare file contains names and addresses of virtually the entire U.S. population 65 years and over. Under suitable circumstances, this file can be made available to U.S. federal agencies for health research studies. If medicare files were used to supplement the sample of blacks and Hispanics, 65 years and over, the screening sample would not be required to provide persons 65 and over, and the 65,000 screeners would be adequate.

2. If the dual-frame sample is not feasible, another option is to combine NHIS data for several years, and produce data for Hispanics 65 and over on the basis of three-year moving averages. Of course, this option can be extended to some of the other rare subdomains. Three years of NHIS would provide a sufficient sample for all subdomains except Hispanic males 65 and over, and it would come close in that subdomain. For NHIS supplements that are obtained in a single year and for which Hispanic data by age and sex are required, it would be necessary to recontact Hispanics in the rarer age-sex subdomains that are in NHIS in the previous year or two.

3. Combined with the options discussed above or independently, modifications can be made in the definitions of the subdomains so they become less rare. An obvious example is to combine data for Hispanic males and females, 65 years and over. Another option under consideration is to reduce the number of age subdomains.

4.3 Person vs. Household Sample

The discussion up to now treated the NHIS as a person rather than a household sample. Sample sizes and sampling rates were developed on the assumption that individuals would be sampled within households to achieve the desired sample sizes. However, experienced survey practitioners are well aware that a substantial part of the data collection costs in household surveys comes from the efforts to contact the sample households. Survey costs are thus much greater when the sample persons are scattered at random in a large number of households than when the sample persons are concentrated in fewer households. As a result, we explored the sample size and variance implications of a sample design in which the sample units are complete households.

We first describe the general procedures that would be followed for this design. For simplicity, our discussion assumes there is no oversampling in areas of high minority concentrations.

1. Start with the rarest subdomains. If there are no modifications in the survey objectives, this subdomain is Hispanic males 65 years and over. The sampling rate for this subdomain is .001712, and this is therefore the screening sample. All households with Hispanic males 65 years and over in the screened sample are automatically included in the NHIS.

Table 5. Oversampling rates in Hispanic density strata for Hispanic subdomains

Subdomain	Design effect*	Sample size	% Hispanic in BG/ED: <5	5-9	10-19	20-49	50+
Hispanic males 65+	1.59	1590	1.5	3	4	5	10
Hispanic females 65+	1.08	1080	1.5	3	3	3	3
Hispanic females 18-24	1.00	1000	1.5	1.6	1.6	1.6	1.7
Hispanic males 18-24	1.00	1000	1.5	1.6	1.6	1.6	1.6
Hispanic females <5	1.00	1000	1.5	1.5	1.5	1.5	1.5
Hispanic males <5	1.00	1000	1.4	1.4	1.4	1.4	1.4
Hispanic males 45-64	1.00	1000	1.2	1.2	1.2	1.2	1.2
Hispanic females 45-64	1.00	1000	1.1	1.1	1.1	1.1	1.1
All other Hispanic subdomains	1.00	xx	.67	.67	.67	.67	.67

*This covers only the design effect from disproportionate sampling, not that due to clustering.

2. Subsample the remaining screened households for retention of households with persons in the next-to-rarest subdomain, that is, Hispanic females 65 years and over. Some persons in this subdomain will have been brought into the sample because they are members of households with the rarest subdomain. In calculating the subsampling rate for the second subdomain, it is necessary to increase the sample size above 1,000 to account for the component of the design effect created by having more than one sampling rate for the subdomain. As before, all persons in this subsample are included in the survey.

3. Continue this procedure subdomain by subdomain, with the subdomains sequenced in order from the rarest to the most common. In each case, calculate the subsampling rate for households not previously selected, taking into consideration the number of expected persons that will be brought into the sample because they live in the same households with persons in rarer subdomains.

4. The procedures described in the three steps above are applied only for the subdomains that are oversampled. The sampling rate for the remaining subdomains is the rate necessary to attain the total number of households described for the survey. For the NHIS, this number is 50,000 households.

Using the data file from the 1988 NHIS, we have analyzed the sampling rates and sample sizes that would result from such a procedure. The sampling rates for each of the 17 subdomains shown in Table 6 are those that would be required for a self-weighting sample, and for the final stratum sampled for each subdomain when a household sample is used. The last column in Table 6 shows the increase in sample size over a self-weighting sample. The increase is due to sample persons being selected at variable rates.

An examination of Table 6 reveals the following.

1. The increases in sample size necessary for most of the subdomains are quite small. The only important increase is for Hispanic females 45-64. Three other subdomains require moderate increases, with trivial increases needed for the rest.

2. Since the sampling rates for the household sample were fixed to produce the same variances for subdomains 1-14 as from the self-weighting person samples, the only difference in variances is for subdomains 15, 16, and 17. With a person sample, a sampling rate of .000478 would be used for these three subdomains. With a household sample, the rate would be .000438. The variances for nonblacks and nonHispanics would therefore increase by a factor which is the ratio of these two rates, or 9 percent. The variances for subdomains 15 and 16 would be lower for a household than person sample because a significant proportion of persons in these two subdomains live in households that will be sampled at higher rates.

3. The number of households to be screened is the same for person and household samples. It is fixed by the sampling rate for the rarest subdomain.

The 9 percent increase in variance for the nonblack, nonHispanic subdomains is the price paid for using a household sample. There would be cost reductions as well as simplification of the field operations because the number of sample persons would be concentrated in fewer households than with a person sample. In our view, it appears to be a useful tradeoff, although of course this is a personal judgment and other analysts may put different values on small losses in precision for the majority U.S. population.

Table 6. Sampling rates with person sample and household sample

Person subdomain		Rate with person sample	Household sample: rate in corresponding household stratum[1]	Percent increase in sample size
Number	Description			
1	M, Hispanic 65+	.001712	.001712	0.0
2	F, Hispanic 65+	.001225	.001106	3.8
3	M, black 65+	.000852	.000852	0.0
4	F, Hispanic 18-24	.000801	.000779	2.4
5	M, Hispanic 18-24	.000775	.000752	1.9
6	F, black <5	.000752	.000751	0.0
7	M, black <5	.000730	.000719	0.2
8	F, Hispanic <5	.000730	.000699	0.8
9	M, Hispanic <5	.000707	.000659	1.7
10	M, black 18-24	.000682	.000662	0.5
11	M, Hispanic 45-64	.000618	.000531	5.7
12	F, black 18-24	.000584	.000481	4.6
13	F, black 65+	.000577	.000506	5.5
14	F, Hispanic 45-64	.000548	.000367	20.8
15	Other Hispanics[2]	.000478	.000438	–
16	Other blacks[3]	.000478	.000438	–
17	Others[4]	.000478	.000438	–

[1] Rate only applies to households with persons in the subdomain excluding those who are in households sampled at higher rates.
[2] Includes Hispanics 5-17 and 25-44, of either sex.
[3] Includes blacks, 5-17 and 25-64, of either sex.
[4] Includes all nonblacks, nonHispanics.

4.4 Number of Sample PSUs

As mentioned earlier, cost functions are not always well-behaved. Determining the optimum number of PSUs for the NHIS is a good example of this. Leaving aside the fact that there is no single optimum solution for all items, another problem arose in attempting to describe the cost as a function of the number of PSUs.

In noncertainty PSUs, most interviewers are responsible for two or three PSUs, in one of which they reside. The NHIS sample is randomly distributed over the 52 weeks of each year, with each noncertainty PSU in the sample from 8 to 12 weeks. Thus, interviewers typically spend about 10 weeks working in the PSU in which they live, and another 15 or 20 weeks in other PSUs. Some of these other PSUs are within commuting distance of the interviewers' residences; other PSUs involve overnight stays with hotel and other per diem costs.

A cost function can be developed that describes how costs will vary when there are small changes in the number of PSUs. However, if the number of PSUs is sharply reduced, for example, from 200 to about 100 or 125, the workload in each noncertainty PSU becomes large enough to keep an interviewer busy during most of the year. At some point it becomes efficient to use only resident interviewers, even if this requires reducing the average workload per interviewer so that the recruiting and supervision costs are higher. With the reduced number of PSUs, a totally different cost function is called for. It is theoretically possible to create a cost function containing dummy variables that take the values of 0 or 1 depending on the number of PSUs. Such a cost function would not be differentiable and could not be solved for the optimum number of PSUs in the usual manner.

There are also problems in understanding what happens to costs when the number of PSUs is increased. Since interviewers in the noncertainty PSUs usually spend about 20 weeks a year travelling to other PSUs and back home, it is reasonable to assume that the travel costs and time would not be much different if the travel during those 20 weeks were to one or two PSUs, or to four or five PSUs. From this perspective, the number of PSUs could be increased dramatically with little impact on costs. The increase in costs would come from a small reduction in the proportion of interviewers' work in home PSUs and in some intangible factors, such as lower efficiency due to interviewers lower familiarity with the road patterns in the PSUs they cover.

However, it turns out that the between-PSU variances for most health items are rather small, typically amounting to 5 to 10 percent of the total variances. Consequently, increasing the number of PSUs by 50 or 100 percent will only reduce the total variances by 2 to 5 percent. It is unlikely that the cost function is accurate enough to detect whether costs would go up that much. A decision on the optimum number of PSUs is thus somewhat uncertain. Luckily, whatever decision is reached will not have an important effect on costs or variances.

5. Conclusions

Our conclusions center around two themes. The first is how better cost data could be collected. The second is on the optimization process.

We were not successful in modeling cost data. Theoretically, the independent variables we used from the survey data should have come much closer to explaining most of the variation in assignment charges. One solution might be to improve the training in keeping track of time and cost and to give more attention to it in supervision. This would obviously have some cost, either in the form of more supervisors and more training hours, or reduced attention to other areas. As we move to various forms of computer-assisted interviewing, certain kinds of information could be made more accessible to the survey researcher. The date of completion will be more accurately and reliable available, as will the administration time. If the survey control system is computerized, interviewers can be required to keep a record of the date, time and outcome of each call to the household on the computer. This could make available a wealth of data at the segment and household level. In the current effort, such data could have allowed us to find the missing link between cost and product by letting us associate the time and expense for a pay period with the actual cases worked on by the interviewer during that same period.

We believe this is a natural way to proceed when a survey has multiple goals and priorities cannot be mechanically established. Regarding the optimization process, it became apparent to us that even with excellent cost data, it would be very difficult to create a cost model. Furthermore, even if such a cost model could be developed, optimizing the design involves more than just using Lagrangian multipliers or numerical methods to minimize the variance on some statistic subject to constrained cost. A survey such as NHIS has multiple and conflicting goals. Some of the goals may be inconsistent with the size of budgets that are likely to be available in the near future.

We first created a sample design that met all the objectives. The costs for the design were extremely high. This led to a modification of survey goals. In effect, an iterative procedure was developed in which the costs of sample designs were examined with particular attention paid to components of cost that might reduce those cost components. When the alternatives turned out not to cut the costs sufficiently, modifications in survey goals were considered. The basic sample design and alternatives were then modified to reflect the new goals, and costs re-examined. The process was repeated for several iterations.

---

[1] G. Kalton and D.W. Anderson, "Sampling Rare Populations," Journal of the Royal Statistical Society, Series A; Vol. 149; 1986.

[2] J. Waksberg, "The Effect of Stratification with Differential Sampling Rates on Attributes of Subsets of the Population," Proceedings of the Social Statistics Section, ASA; 1973.