VARIANCE ESTIMATION CONCERNS IN THE NATIONAL MEDICAL EXPENDITURE SURVEY (NMES) SCREENER SAMPLE

David J. Fitch and Steven B. Cohen, National Center for Health Services Research
David J. Fitch, 5600 Fishers Lane, Rockville, MD 20857

Introduction

The Household Survey Component (HHS) of the National Medical Expenditure Survey (NMES) was established to provide an assessment of the health care utilization, costs, sources of payment, and health insurance coverage that characterize the U.S. civilian, noninstitutional population. The survey was designed to provide data for a major research effort in the Division of Intramural Research of the National Center for Health Services Research and Health Care Technology Assessment (NCHSR), and was cosponsored with the Health Care Finance Administration (HCFA). The data will meet the needs of government agencies, legislative bodies, and health professionals for more comprehensive data required for the analysis and formulation of national health policies.

The NMES household sample includes oversampling of groups of particular policy interest: blacks, Hispanics, the poor and near poor, the elderly and persons with functional impairments. This was accomplished by conducting a separate screening interview in the fall of 1986 to facilitate the oversampling of the policy relevant subgroups for the NMES. The NMES household survey sample design can be characterized as a stratified multi-stage area probability design. The survey was conducted from two independently drawn national samples selected from two distinct survey organizations, Westat, Inc. and NORC. Data collection from households for calendar year 1987 was conducted from January 1987 to July 1988.

The purpose of this paper is to measure the precision of survey estimates that characterize the NMES Screener sample. The paper will include a summary of the adopted sample design, and raise questions regarding the consideration of alternative sample design strategies that would further improve upon the precision of survey estimates. The paper includes an analysis that compares two methods of variance estimation that are appropriate for application to NMES data, the balanced half-sample method, and the Taylor series linearization method. The methods are compared with respect to estimated standard errors, user facility and computation cost.

Survey Design Considerations

Data from complex surveys require special consideration with regard to variance estimation and analysis, as a consequence of design departures from simple random sampling. The design effect measures the impact of a given survey design on the precision of survey estimates, relative to a design with simple random sampling assumptions. The design effect is defined as the ratio of the variance of a survey estimate under a given design, relative to the variance of the same survey estimate under simple random sampling assumptions.

The design effects obtained are influenced by two fundamental design considerations in national surveys, the application of each being limited by cost constraints. The first consideration is that the population is to be divided into strata homogeneous with respect to the variables of interest. The second consideration is that, within each such strata, units are to be grouped into heterogeneous clusters, with the clusters being the units that are sampled. In practice, when selecting a sample of U.S. households, counties are defined as the units in the first stage, grouped into strata. Within each such stratum, counties are then selected.

Selecting a county as a primary sampling unit (PSU) has the very large advantage of being a geographical area conveniently worked by one interviewer or one team of interviewers. In theory, one could specify strata as being all blocks or enumeration districts with median monthly rent or median home value, within specified ranges. This would typically result in more homogeneous strata, as the first consideration specifies, but would result in widely scattered households, adding a cost inefficiency to the interviewing process. With the usual procedure where strata are represented by sampled counties, considerable heterogeneity within strata prevails. While this strategy does not result in a optimal design under the first consideration of obtaining homogeneous strata, it allows for the possibility of very large gains from the application of the second consideration.

The NMES Sample Design

An examination of the analytical goals of the study and budget constraints indicated that the sample design for the NMES household survey should satisfy the following requirements:

1. The full series of interviews should be completed in approximately 14,000 households.

2. The sample should be spread over at least 100 separate areas to represent the civilian non-institutionalized population of the 50 states and the District of Columbia.

3. The sample shall be designed to produce unbiased national estimates and unbiased estimates with adequate precision for the four census regions.

4. The sample shall be designed to meet predesignated precision specifications for the following population subgroups of analytical interest: blacks, Hispanics, the poor and near poor, the elderly (65+), and the functionally impaired. It was recognized that in order to achieve the NMES precision requirements for the overall population and specified domains of interest, an initial

sample selection substantially larger than 14,000 households with subsequent screening would be required.

The numerical constraint on a minimum number of geographical localities was specified to ensure sufficient geographic dispersion of the sample and allow for separate regional estimates. The precision specifications were provided to insure that the design would meet analytical objectives and to facilitate stage-specific sample size determination.

The adopted NMES household survey sample design is a stratified area probability design with four stages of sample selection: (1) selection of primary sampling units (PSU's), which are counties, parts of counties or groups of contiguous counties; (2) selection of segments within PSU's; (3) selection and screening of dwelling units within segments; and (4) selection of dwelling units based on demographic characteristics (both household and individual level) from the set of screened dwelling units.

The sample of PSU's represents a union of the national sample frames of Westat, Inc., and NORC. The national general purpose area samples from Westat and NORC that comprise the NMES household sample are similar in structure, thus simplifying the development of sampling, listing and interview procedures and permitting a single management control system. Since the sampled PSU's, segments, and dwelling units selected for screening were representative of two independently drawn national samples, except for difficulties associated with survey nonresponse and other non-sampling errors, statistically unbiased national and domain specific estimates can be produced from each sample or from the two samples combined. The combined sample includes 165 PSU's located in 127 distinct sites. The number of separate primary areas is less than the total primary sampling units in the two national samples because some areas are in both samples.

More specifically, the Westat first stage sample was stratified by social, economic and demographic characteristics which included region, SMSA status, percent of population employed, percent white, and percent over age 65. In all, 81 PSU's were selected for NMES. Similarly, the NORC first stage sample included the following stratification measures: region, SMSA status, and population size. The NMES sample consisted of 84 NORC PSU's. Within PSU's, a two or three stage sample design was used to select dwelling units for the screening sample. The first stage consisted of 1980 Census Enumeration Districts (ED's) or individual block or block combinations. The second stage was only used when ED's or block groups were exceptionally large in area or number of households, respectively. Then "chunking" or the partitioning of the ED or block group was employed, dividing the selected area into several smaller segments of approximately equal size in terms of households, one of which was randomly selected. The third stage consists of the selection of all or a systematically selected subsample of households within each area segment. The ED's or blocks were selected with probability proportionate to size, with a systematic procedure allowing for implicit geographic stratification. The sampling and subsampling rates were specified so that all dwelling units in the U.S. had an equal probability of selection.

Within the sampled PSU's, 2,317 segments were selected (1,150 for Westat, 1,167 for NORC). The segment sampling process resulted in a set of maps showing the boundaries of the sampled segments and their associated probabilities of selection. The addresses within the boundaries of sample segments were then listed by trained interviewers, and served as the sampling frame from which the address sample for the NMES screener interview was selected. Approximately 35,000 addresses were selected for screening, within the sampled segments. Following the screening interview a subsample of dwelling units were selected for the full panel household survey from those screened, according to person and household level demographic characteristics. Subsampling rates were specified to obtain the required sample size to satisfy NMES precision specifications for person level estimates.

NMES Screener Sample

As indicated, the 1987 NMES household survey was designed to provide an assessment of the annual health care utilization, expenditures, sources of payment and health insurance coverage in 1987, for the civilian non-institutionalized U.S. population. In addition, the NMES specifically targeted population subgroups of special policy interest for oversampling. More specifically, the NMES design required selective oversampling of blacks, Hispanics, the poor and near poor, those 65 years of age or older, and the functionally impaired.

The poverty status classification was based on the Poverty Index, developed at the Social Security Administration in 1964 and revised by the Federal Interagency Committees in 1969 and 1980. The poverty thresholds are updated every year to reflect changes in the Consumer Price Index. The poverty status classification is defined at the family level (or person level for unrelated individuals) and is a function of family size and, for single or two person families, age of household head. The near-poor classification was defined to consist of families with family income above the poverty level, but less than 1.25 times the poverty level. The 1986 family incomes and poverty level thresholds were used for classification purposes. To facilitate the acquisition of relevant income data to allow for the poverty status classification, show cards were administered during the screener interview. A set of show cards were printed to reflect the different income thresholds for a poverty status classification of poor, near-poor, or other income. The interviewer would then present the appropriate showcard to the household respondent, based on family size, to ascertain the applicable poverty status classification for the household.

The functionally impaired are defined as persons needing or receiving personal assistance or the help of special equipment for a period of three months or more to perform one or more specifically defined activities of daily living (eating, transferring from bed or chair, dressing, bathing,) because of a health or physical problem.

A set of strict precision requirements were specified to allow for detailed analyses of these policy relevant population subgroups (Cohen et al., 1987). The NMES requirement of an average design effect of 1.7 for survey estimates was specified to insure an efficient survey design. To satisfy this requirement, an average segment size of 6 households per segment was planned for NMES. Given the overall sample size requirement of 14,000 household interviews, 2,317 segments were selected, nearly twice the number considered in the 1977 National Medical Care Expenditure Survey where there were 1290 segments sampled (Cohen, 1983).

Due to the complex sampling requirements in the NMES, a separate screening interview was used to facilitate sample identification of population subgroups targeted for sampling at differential selection rates. The separate screening operation allowed for analyses of the results of the screener interview prior to the selection of the NMES Round 1 household sample. More specifically, it allowed for analysis of three features of the sample before the design was finalized: (1) a determination of whether the number of households screened was sufficient to provide the required sample sizes for each analytical domain; (2) whether the person based domain requirements were consistent with a sample size specification of 14,000 households completing the full series of interviews; and (3) development of the subsampling rates needed to attain the desired sample size in each domain. An advanced screening operation would make available precise counts of the outcome of the NMES subsampling process, prior to fielding the Round 1 household sample.

Variance Estimation Methods

In order to derive a population estimate of a health care measure which is representative of the NMES data set, sampling weights must be considered that account for differential selection probabilities, nonresponse, and poststratification. Since the sampling weights are initially defined as the reciprocal of an individual's probability of selection, the weight that would be applied in NMES for oversampled population subgroups (e.g., the elderly), would be smaller than weights that represent the remaining sample population subgroups. Nonresponse adjustments are the applied to sampling weights to reduce the error in survey estimates associated with nonresponse bias.

With respect to nonresponse, although no person level information may be available to characterize the nonrespondents, information at the segment level or PSU level is often available on aggregate population characteristics. Weighting classes would then be formed using this information, and the sampling weights of all survey respondents within a weighting class would be inflated by an adjustment factor to represent the sum of the sampling weights of all the original sample represented in the weighting class. Alternatively, the core NMES household data could be missing for a subset of individuals who were screener respondents. Weight adjustments would then be made within weighting classes defined by person specific characteristics (e.g., age, sex and poverty status)

Had there been no sampling error, the weights for each age-race-sex population subgroup in the NMES screener sample would have summed to the number of individuals in that age-race-sex category for the nation. With sampling error and with some of the weight adjustments introduced to correct for missing responses, this clearly will not be the case. A poststratification to the NMES screener weights will correct the NMES population estimate for a population subgroup, and adjust the total to a more accurate secondary source. In NMES, the screener sampling weights were adjusted within age-race-sex categories, to sum to Census Bureau estimates for the category, derived from the November 1986 Current Population Survey.

Population estimates of health care parameters produced from the NMES Household Survey, and of demographic measures derived from the NMES screener survey, are generally specified as ratio estimates. If every sample member had the same weight, or if every sample member in each stratum had the same weight, a straight forward linear variance estimator could be used. This was not the case for the NMES screener sample.
In our NMES applications, we considered two of the most widely used variance estimation procedures appropriate for complex survey designs, to obtain an estimate of the precision of our survey estimates. More specifically, we considered the Taylor series linearization method and the method of balanced repeated replication (Wolter, 1985).

The method of balanced repeated replication required that the full screener sample be divided into an orthogonally balanced set of half samples, 104 for NMES, according to rather strict requirements. This technique called for the same adjustment and estimation procedures to be applied for each half sample, as was specified for the full sample. Consequently, the same corrections for non-response and missing data, the same imputation procedures, and the same poststratification adjustments adopted for the full screener sample had to be replicated for each half-sample. If these half samples are contructed according to specific sampling theory (Wolter, 1985), unbiased variance estimates can be obtained with k half samples where k is a number divisible by 4 and which is equal to or greater than the number of strata (k=104 for

NMES).

Should it not be apparent, the construction of 104 sets of weights, each set with a weight for each of the 76,450 observations in the sample (observations not in a half sample having a weight of 0), requires a significant amount of computer time. The weight construction computations for the 104 half samples was done in SAS, and used about seven minutes of computer CPU time on the IBM 3090 system at the National Institutes of Health at a cost of about $300. It is important to note that replicate weight construction is a one time effort. Once constructed, this same matrix of weights is used for any variance estimate using the balanced half-sample method with NMES screener data.

The mathematics, and hence the programming, to apply the Taylor series is much more complicated than that for the half-sample method. The later would more likely be the choice if one had to independently program each technique. Fortunately, well specified software is available for application of the linearization method to complex survey data (Cohen, Burt and Jones, 1986). The program that was considered for the NMES screener data was SESUDAAN (Shah, 1981), which is written in FORTRAN as a SAS procedure. For a variance estimate more fully reflecting the sample design (i.e., households within segments within PSUs within strata), the program requires stage specific counts for each observation of the sample number of elements selected and the corresponding population total in the frame.

The WESVAR procedure (Westat, 1988) was used for the balanced half-sample runs, which was also programmed in FORTRAN as a SAS procedure. The procedure is user friendly, straight forward in its application and well documented. The balanced half-sample method has an attraction which is particularly appealing. One would expect that most variance estimation programs would produce a variance estimate for a mean and/or a total. However, there are occasions when a variance estimate for some other statistic is required, and this can readily be obtained with the half-sample method. As a consequence of the method's straight forward specification, the statistic of interest (mean, total, correlation, etc.) is separately computed using the data in each half-sample, and also using data for the full sample. A variance estimate is then computed from this set of half sample statistics and the one full sample estimate.

Estimated Standard Errors and Design Effects for NMES Screener Data

For the purposes of this study, a representative set of demographic measures which characterized the NMES Screener respondents were selected as criterion variables. The complete NMES screener database contains observations on 76,450 individuals. The following seven demographic measures were included in the variance estimation comparisons: age, gender, veteran status (veteran/other), race (white/non-white), marital status (married/other), functional limitation (any functional limitation/ no functional limitation), and the number of functional limitations.

Standard errors were derived for each of the seven NMES screener demographic measures using the SESUDAAN and WESVAR procedures.
There was a considerable difference observed in the computation time and cost associated with running the respective procedures. Using the IBM 3090 system at the National Institutes of Health (NIH), the half-sample method (WESVAR) required 71 seconds of CPU time to produce the seven estimated population means and standard errors for the specified measures, at a cost of $42.42. Alternatively, the linearization method required only 44 seconds of CPU time to generate the required estimates, at a cost of $26.06.

In theory, it appears these half sample computations could be done at NIH with much greater efficiency, requiring significantly less CPU time. These half-sample computations can readily be re-expressed in matrix form. Consider the following scenario. A person by half-sample matrix of weights, in this case a 76,450 by 105 matrix, is premultiplied by a 7 by 76,450 matrix - a variable by person matrix. This gives a 7 by 105 matrix of estimates from which the variance estimates are computed in an efficient fashion. NIH has installed vector hardware into their IBM 3090 computing equipment, facilitating much faster matrix multiplication computations.

Table 1 presents a comparison of the estimated standard errors and related design effects for mean estimates of the seven demographic measures selected from the NMES screener database, using the alternative methods of variance estimation. Standard error estimates and related design effects were similar across the two methods for mean estimates of (1) the number of functional impairments, (2) the proportion of the population with an impairment, (3) the proportion of the population that are veterans, and (4) the proportion of the population that is married. This was determined by computing the ratio of design effects for the two methods of variance estimation and testing for the equivalence of the ratio to equality. The ratio of design effects is equivalent to the ratio of estimated variances. Consequently, an F test at =.05 with 104 degrees of freedom (number of strata) was used to determine whether the variances derived by the two techniques was statistically equivalent.

The balanced half-sample method of variance estimation clearly yielded more precise estimates for mean estimates of (1) age, and (2) gender (proportion of the population that are male). This is a consequence of the replication method accounting for the effect of poststratification on precision, which is not being captured by the Taylor series linearization method. Although the poststratification adjustment considered race, the race measure that was considered in our analysis (white/nonwhite) departed from the poststratification classification (Hispanic, black-non-Hispanic, other). Again, the balanced half-sample method of variance estimation yielded a more precise variance estimate than the Taylor series method on a variable associated with poststratification.

Generally, the design effects for the NMES screener survey were well behaved, varying between 2.4 and 4.6 for the mean estimates of demographic measures that were not associated with poststratification. The high design effect observed for that national estimate of the population that was white was significantly greater than expected (design effect >10). This was potentially a function of the large average segment size that characterized the NMES Screener sample (12 dwelling units per segment).

For the purposes of comparison, design effects are presented (Taylor series approximation) for the same mean estimates of the selected demographic measures, with the sample constrained to only include the head of the primary household in each dwelling unit. This analysis attempted to discern the impact on survey design effects when the effect of clustering within households is removed. The result of this restriction was greater stability in survey design effects, and a general decline in their magnitude.

Summary

This paper has concentrated on variance estimation concerns with respect to the NMES Screener Sample. The sample design is described in detail, to convey the complex survey design that was adopted with its significant departures from simple random sampling assumptions. In addition, the precision of survey estimates was examined for a representative set of demographic measures selected from the screener database. A comparison of two methods of variance estimation appropriate for complex survey data was also considered, which indicated general convergence of variance estimates across methods. The study demonstrated, however, that the balanced half sample method best controlled for the effects of poststratification, but the Taylor series method was less costly to apply.

We expect to see lower design effects for estimates derived from the NMES household survey. This is a consequence of the large number of segments that were included, and the reduction in the average segment size to six households per segment in the household survey. Further analyses will be considered using the NMES household data, to suggest redesign strategies in future surveys that will improve the precision of survey estimates.

References

Cohen,S.B. (1983). Design Effect Variation in the National Medical Care Expenditure Survey. American Statistical Association, Proceedings of the Survey Research Section.

Cohen S.B.,Burt,V.B.,and Jones,G.K. (1986). Efficiencies in Variance Estimation for Complex Survey Data. The American Statistician, 40,No.2,157-164.

Cohen,S.B., DiGaetano,R., and Waksberg, R. (1987). Sample Design of the National Medical Expenditure Survey - Household component. American Statistical Association, Proceedings of the Survey Research Section.

Shah,B.V. (1981). SESUDAAN:Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. Research Triangle Institute. Research Triangle Park,NC.

Westat,Inc. (1988). The WESVAR Procedure. Rockville,MD.

Wolter,K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

TABLE 1

ESTIMATED STANDARD ERRORS (S.E.) AND DESIGN EFFECTS (D.E.) FOR MEAN ESTIMATES OF SELECT DEMOGRAPHIC MEASURES FROM THE NMES SCREENER DATABASE.

	Method of Variance Estimation		
	Half Sample	Linearization	
	n= 76,450	n= 76,450	DU Head only
No. of functional limitations			
S.E.	.0056	.0061	.0106
D.E.	4.38	5.14	3.95
Age			
S.E.	.0341	.199	.221
D.E.	.18	6.30	4.52
Any functional limitation			
S.E.	.0016	.0018	.0032
D.E.	4.61	5.29	4.14
Served in armed forces			
S.E.	.0018	.0021	.0051
D.E.	3.06	3.18	3.66
Married			
S.E.	.0028	.0033	.0058
D.E.	2.36	3.54	3.94
Sex			
S.E.	.0000	.0015	.0061
D.E.	indeterminate	.69	4.84
White-nonwhite			
S.E.	.0054	.0087	.0075
D.E.	10+	10+	10+

Source: NMES Screener Sample, National Center for Health Services Research and Health Care Technology Assessment (NCHSR).