Distribution and Properties of Variance Estimators for Complex Multistage Probability Samples

An Empirical Distribution

This report presents results from an empirical investigation of the behavior of the replication and linearization techniques for measuring variance. The study utilizes data collected in the Health Interview Survey. Most scientific sample surveys are based on complex multistage probability designs, with design components that include unequal probabilities of selection of elements in the population, stratification, and several stages of clustering. The estimation procedures usually involve nonresponse and poststratification adjustments. These actions cause concern in the methods of estimation of standard errors and their subsequent use in constructing confidence intervals and testing hypotheses. There are several ways to study the characteristics of variance estimates. The method in this report is an empirical investigation. Data that were collected in a national sample survey become the universe and repeated samples are drawn from it. For each sample the statistics under study were calculated and sampling distributions of the estimates were generated. The material in this report was taken, in part, from Dr. Bean's doctoral dissertation.

DHEW Publication No. (HRA) 75-1339

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

Health Resources Administration
National Center for Health Statistics
Rockville, Md. March 1975

Library of Congress Cataloging in Publication Data

Bean, Judy A.

Distribution and properties of variance estimators for complex multistage probability samples.

(Series 2: Data evaluation and methods research, no. 65) (DHEW publication no. (HRA) 75-1339)

"Final report."

Supt. of Docs. no.: HE 20.2210:2/65

1. Sampling (Statistics) 2. Estimation theory. 3. Monte Carlo method. I. Title. II. Series: United States. National Center for Health Statistics. Vital and health statistics. Series 2: Data evaluation and methods research no. 65. III. Series: United States. Dept. of Health, Education, and Welfare. [DNLM: 1. Population surveillance. 2. Probability. 3. Sociometric technics. W2A N148vb no. 65]

RA409.U45 no. 65 [QA276.6] 312'.07'23s [519.5'2]

ISBN 0-8406-0029-1 74-16356

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $1.10

Vital and Health Statistics-Series 2-No. 65

DHEW Publication No. (HRA) 75-1339

Library of Congress Card Number HRA 74-16356

CONTENTS

CONTENTS—Con.

SYMBOLS

Data not available	---
Category not applicable	...
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

DISTRIBUTION AND PROPERTIES OF VARIANCE ESTIMATORS FOR COMPLEX MULTISTAGE PROBABILITY SAMPLES

Judy A. Bean, Ph.D., *University of Iowa*

INTRODUCTION

Background

Since the early 1940's, there has been a substantial growth in the use of surveys of human populations, which has occurred in all research areas, including the biological and health sciences, and particularly the social sciences. Along with the increasing utilization of surveys, the purposes for which data are collected have expanded. Research workers first employed surveys to obtain specific information about groups for descriptive purposes only. Increasingly, however, they have become more interested in making comparisons among subgroups of the population, in testing hypotheses about the population, or in disclosing complex relationships. Both the greater use and the change in objectives have promoted considerable theoretical and practical development in sample design which, in turn, has resulted in decreased use of the traditional method of random selection from human populations.

Most scientific sample surveys today are based on complex multistage probability samples, with design components often including unequal probabilities of selection for different elements in the population, stratification, and two or more stages of clustering. The estimation procedure may involve adjustment for nonresponse, use of concomitant variables, and poststratification. Because of these features, the assumptions of independence and equal probability of selection are not valid. Therefore, analytical questions arise. Theory has not been developed sufficiently to overcome all the difficulties evoked by the correlation induced by clusters of sample units.

One analytical problem not solved in theory concerns estimation of variances and standard errors of parameter estimates. Since these are integral components of the formulas for constructing confidence intervals and testing hypotheses, standard errors are crucial in statistical inference. When variance estimates are needed for statistical inference, several choices are available.

There is the option of designing and carrying out a random sample but, as mentioned previously, this has become an uncommon design. The human populations to be studied today are large and widely dispersed, which makes listing and travel prohibitively expensive. However, it is frequently possible to use a reasonable approximation for estimating variances when the exact formula is not known. Care must be exercised when selecting an adequate approximation. For instance, if the random sample formula PQ/n is substituted for the appropriate variance estimator for clustered samples, it can be shown that the variance is usually underestimated.

The investigator can perhaps translate the problem for which no estimates of standard errors are obtainable into one for which estimates are available. For example, the 2 × 2 chi-square test can sometimes be translated into the difference between two proportions. Inferences can be made on statistics for which the true variances cannot be calculated, by applying variances of similar statistics. For clarification, consider the situation in which an analyst is interested in testing whether or not k ($k > 2$) means are from the same population. Since the variance for the difference of two means is known, the variance for several pairs or perhaps all pairs can be computed. Then the ratio of these variances to variances that would have been obtained if the design had been a simple random sample can be calculated. The analyst then infers that the comparison of the correct variance estimate of the k means to simple random sample variance would yield a similar ratio. Thus, he can use this ratio as an adjustment factor to the usual F ratio for testing such a hypothesis.[1,2]

Another approach to the problem of estimating variances is to interweave within a sample design a small number of replications. Each of these replications produces an estimate of the parameter. Therefore, their comparison provides an estimate of variance for the sample. An example of this is to draw 10 independent subsamples from the same population using the same probability design. Let $\bar{x}_i$ denote the estimated mean of the ith subsample and

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} \bar{x}_i$$

Then from the 10 independent estimates, the simple variance of the mean $\bar{x}$ for the entire sample can be computed. This estimate is

$$\hat{\text{VAR}}(\bar{x}) = \frac{1}{10}\frac{1}{9}\sum_{i=1}^{10} (x_i - \bar{x})^2$$

This method is too expensive to employ for a complex design.[3]

A similar procedure to using independent subsamples is the random group method. To illustrate the technique, suppose one has a simple random sample of n observations drawn with replacement. The observations are randomly distributed among t groups consisting of n/t observations each. The variance estimate is

$$\hat{\text{VAR}}(\bar{x}) = \frac{1}{t}\frac{1}{t-1}\sum_{j=1}^{t} (\bar{x}_j - \bar{x})^2$$

where

$\bar{x}_j$ = the mean for the jth random group, and

$$\bar{x} = \frac{1}{t}\sum_{j=1}^{t} \bar{x}_j.$$

This method is suitable for multistage survey samples. If primary sampling units (PSU's) are drawn and then second-stage units are sampled within, all the second-stage units from a primary unit constitute a single unit when forming the random groups. If there are few PSU's, this procedure is not too useful. Moreover, there is a loss of information.[4]

The final choice is to compute the estimates of standard errors by one of the three general methods: Taylor series expansion or linearization method, pseudoreplication or replication method (originated from the methods of independent replication and random group), or jackknife method. Unfortunately, the behavior of the variance estimates produced by these methods is not thoroughly understood.

Three major survey organizations—the U.S. Bureau of the Census; the National Center for Health Statistics (NCHS); and the Survey Research Center (SRC), University of Michigan—commonly use one of the three general methods, linearization, replication, or jackknifing, to calculate estimates of variances from the sample data. It is, therefore, important to become familiar with the properties of these methods.

Strategy of the Study

There are several possible ways to study the characteristics of variance estimates created by the general methods. One possibility is the development of applicable theory by assuming a certain distributional form and obtaining exact analytical solutions. However, the mathematics involved have so far proven to be intractable.

There is another drawback with this approach with a stratified multistage design: two sources of sampling error are between the first-stage units and within the first-stage units. When only one unit is chosen from a stratum, there is no consistent way of estimating the variance between the first-stage units from the sample itself. In practice, the designs seldom fulfill this simple assumption of the selection of two units. Another avenue of study is Monte Carlo sampling from synthetic populations. Such populations are difficult to construct and are of questionable representativeness. A third device is an empirical investigation. By this approach, data that have been collected in a survey become the universe, and repeated samples are drawn from it. For each sample, the statistics being studied are computed, and sampling distributions of the estimates are generated. Perhaps the most famous such study is the investigation by "Student" in 1908,[5] which resulted in his derivation of the *t* distribution.

In this report, the view is that by Monte Carlo sampling of data collected in a nationwide sample, variance estimators can be studied to gain insight into their properties and behavior. An empirical investigation of the behavior of the replication and linearization variance estimation methods from a viewpoint that is both broad and practical is undertaken. The study uses five variables collected on 131,575 people in the U.S. Health Interview Survey (HIS). The main objectives of the research were the following:

1. To investigate two general variance estimator methods,[a] linearization and replication
2. To study the distribution of the ratio of an estimated mean minus its expected value divided by its standard error
3. To measure the impact of poststratification on variance estimates

Other aspects examined in the investigation were the extent of the biasness of two estimators and the feasibility of using a simpler variance estimator as an approximation to the correct replication and linearization variance estimates (the results are given in appendixes I and II).

When a study of this magnitude is described and discussed, there is the danger that the reader will lose sight of the overall objectives amidst all the details of the analysis. To facilitate reporting, the main features of the methodology involved in doing the research are listed, as follows:

1. *Sample design.* By this term is meant the components of selection of the sample units. The design used includes stratification, sampling of the first-stage units with probability proportional to size, the largest units entering the sample with probability 1, and finally, subsampling clusters of households within the first-stage units, thereby embedding two stages of clustering in the design.
2. *Estimation procedure.* This refers to the method for producing estimates of the health characteristics. Two different poststratified estimators of means are used.
3. *Variance calculations.* This feature is the computation from the sample data of estimates of the variance of the estimators produced in feature number 2. For each estimate of a characteristic, the variance is estimated in a number of different ways.

It is appropriate to state the reasons why only two out of the three general variance estimators are studied here. The intent of the research is to observe the variance estimators produced by the methods in more depth than has been done previously and to study the behavior of the estimators for types of designs and estimation procedures being used today. Data handling, even for a small-scale study of a similar nature, is quite massive, and this problem is compounded by the type of design and procedures selected. Therefore, the decision was made that only two of the three variance estimator techniques would be investigated.

HIGHLIGHTS OF THE FINDINGS

This study was undertaken to investigate how well statistical methods important for making

[a]Since the replication and linearization schemes are actually being used by several samplers, the author chose to study only these two techniques in this study.

inferences are valid when the data have been collected in a multistage probability sample survey and subjected to a complex estimation procedure. The main points examined were the following:

1. The behavior of the variance estimates produced by the balanced half-sample replication method and the linearization method
2. The distribution of the ratio of a sample estimate minus its expected value to its estimated standard error
3. The impact of the estimation technique poststratification on variance estimates

Since the mathematics have not been developed to answer this question of validity, the approach was to use Monte Carlo simulation from a specified universe and to determine the empirical results.

The sample design was a stratified two-stage cluster design with first-stage units being selected with probability proportional to size. The estimation process involves weighting by the reciprocal of probability of selection and poststratification. Five variables were used, and three different samples sizes were studied with 900 samples of each size drawn.

Because design III is the largest sample size and, thus, the one most comparable to real surveys, the highlights for only this design are presented. Characteristics of the universe and a typical sample are:

	Universe	*Average sample*
Number of PSU's	149	30
Number of segments	7,768	600
Number of persons	131,175	8,772 (approximate)

Table 1 gives the major results for the variance of a ratio estimate produced by the replication method, VAR($R1S$) and the linearization method, VAR($L1$).

The proportion of times the standardized variable fell within certain regions was calculated for each of the five variables. These proportions were averaged across the five variables and compared with the expected proportions of a t distribution with 19 degrees of freedom, and the normal distribution, in table 2.

Table 1. A summary of the major findings for design III

Finding	Variable: Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing physician
Population parameter R	8400.0	14.6716	4.6758	1.0557	0.6842
Average sample estimate R_1	8392.8	14.6595	4.6548	1.0597	0.6840
Sample variance of $R_1 - s^2_{R_1}$	27,445.7	0.9496	0.0168	7.0474×10^{-3}	7.8942×10^{-5}
Average value of the variance estimate VAR($R1S$)	26,554.8	0.9206	0.0178	7.0842×10^{-3}	8.3010×10^{-5}
Bias of VAR($R1S$)	−890.9	−0.0290	0.0010	0.0368×10^{-3}	0.4068×10^{-5}
Variance of VAR($R1S$)	$10{,}360.0 \times 10^4$	1.1370	0.8132×10^{-4}	0.0129×10^{-3}	0.1089×10^{-8}
Relative bias of VAR($R1S$)	−0.0325	−0.0305	0.0595	0.0052	0.0515
Relative variance of VAR($R1S$)	0.1375	0.1519	0.2882	0.2596	0.1747
Average value of the variance estimate VAR($L1$)	26,174.7	0.8915	0.0175	6.7380×10^{-3}	8.1135×10^{-5}
Bias of VAR($L1$)	−1,271.0	−0.0581	0.0007	-0.3094×10^{-3}	0.2193×10^{-5}
Variance of VAR($L1$)	$10{,}480 \times 10^4$	0.1310	0.7959×10^{-4}	0.0113×10^{-3}	0.1086×10^{-8}
Relative bias of VAR($L1$)	−0.0463	−0.0612	0.0417	−0.0439	0.0278
Relative variance of VAR($L1$)	0.1391	0.1453	0.2820	0.2275	0.1743
Relative bias of VAR($R1S$) / Relative bias of VAR($L1$)	0.7019	0.4984	1.4269	−0.1185	1.8525
Relative variance of VAR($R1S$) / Relative variance of VAR($L1$)	0.9885	1.0454	1.0220	1.1411	1.0023

Table 2. Distributions of $\frac{R_1 - E(R_1)}{\text{VAR}(R1S)}$ and $\frac{R_1 - E(R_1)}{\text{VAR}(L1)}$ compared with expected values under t_{19df} and normal[1]

Standardized deviate	Normal	t_{19df}	Average across variables: VAR($R1S$)	Average across variables: VAR($L1$)
$\pm 1\sigma$	0.683	—	0.675	0.668
$\pm 1.64\sigma$	0.900	0.879	0.878	0.872
$\pm 1.96\sigma$	0.950	0.932	0.928	0.925
$\pm 2.57\sigma$	0.990	0.976	0.977	0.977

[1]The hypothesized degrees of freedom are the number of strata 19.

REVIEW OF LITERATURE

Replication Method

Although, in the past three decades, methods for selection of sampling units and procedures for producing estimates from survey data have become sophisticated, only in relatively recent years have the methods of analyzing data for such complex surveys begun to be considered. The first articles in the literature on the topic described and discussed a variety of techniques: random groups, replicated samples, interpenetrating samples, and duplicate samples, which are the forerunners of the general replication method as applied in this study. Since this investigation is concerned with the technique in its present-day form, the previous methods will not be reviewed here, but their essential features are described elsewhere.[6]

Before giving a brief account of the development and evidence of the reliability of the replication estimator of variance, its main features are outlined. The basic premise of the method is a very simple one. The estimator X' of a population parameter is calculated from the entire body of sample data (parent sample). Another estimator Y' of the same population parameter is made, using only the data from half the sample, randomly selected. The quantity $(Y' - X')^2$ is an estimate of the variance based on one-half sample (replicate). However, this estimate, depending on only one difference term, has a high variance; repeated differences are required to produce a stable variance estimate. The estimator of variance is simply the mean of these half-sample estimates.

$$\text{VAR}(X') = \frac{1}{k}\sum_{\alpha=1}^{k}(Y'_\alpha - X')^2$$

where k is the number of replicates used. There are three other rather similar forms of variance estimators.

1. $\frac{1}{k}\sum_{\alpha=1}^{k}(Y''_\alpha - X')^2$

where Y''_α is the estimator of the parameter made from the complementary set of data for the αth replication (the half of data not used in the Y'_α estimator).

2. $\frac{1}{2}\left[\frac{1}{k}\sum_{\alpha=1}^{k}(Y'_\alpha - X')^2 + \frac{1}{k}\sum_{\alpha=1}^{k}(Y''_\alpha - X')^2\right]$

3. $\frac{1}{4k}\sum_{\alpha=1}^{k}(Y'_\alpha - Y''_\alpha)^2$

The U.S. Bureau of the Census[7,8] was the first major survey organization to employ the method. The technique was used to estimate sampling errors of ratio estimates produced from the Current Population Survey (CPS) from 1954 through 1964.[b] Gurney[9] worked on the theoretical development of the method for the U.S. Bureau of the Census.

The National Center for Health Statistics draws inferences from data gathered in complex probability sample surveys. The Center uses the replication method to provide variance estimates for the estimates produced from the Health Examination Survey (HES), which consists of a direct physical examination of a probability sample of approximately 7,000 noninstitutional U.S. civilians.[10] Its purposes are to make prevalence estimates for certain medical and dental conditions and to determine the distributions of many physiological characteristics. The survey is conducted in cycles with different age groups sampled for each cycle.

[b]CPS is a monthly nationwide survey of sample households conducted to provide measurements of the labor force. National estimates of employment, unemployment, and other labor force characteristics are made.

The sample design of the survey includes the components stratification, clustering, and two stages of sampling. Along with this highly sophisticated sample design, an equally refined estimation process is used. Features of the procedure are simple inflation of the basic data collected on each sample unit by the reciprocal of the probability of selecting that unit, adjustment for nonresponse, poststratification, and possibly another adjustment factor.

The National Center for Health Statistics, in its efforts to provide the best estimate of variance for this ratio estimate, has used several versions of the replication technique and, from this experience, has developed a routine computer program to provide the analyst with prevalence estimates and their variances. The analyst, by simply requesting the estimates for the total or specified subgroups of the population, receives an estimate of the aggregate value of the health characteristic (numerator), an estimate of the population of the total or subgroup (denominator), and the ratio of the two values. In addition, the analyst obtains an estimate of the standard errors for these three quantities. A more detailed description of the development and use of the replication method can be found elsewhere.[6,11,12] Based on research conducted by NCHS and the performance of the replication variance estimates, the technique appears not only appropriate for the type of estimate produced in its surveys but also useful for variance estimates of other statistics such as regression coefficients, multiple correlation coefficients, and partial correlation coefficients; moreover, the technique has potential for being employed in other forms of statistical analysis. For instance, replication variance estimates are used in a modified sign test and in the calculation of a pseudo-chi-square statistic, which provides a test of independence in a two-way table.

Valuable theoretical work has been done on the replication method.[6] If a subset of all the possible half-samples that could be formed is selected randomly, the number necessary to produce a stable variance estimate is large. By the use of a stratified design with two independent selections made from each stratum and an estimate of the mean, it was demonstrated that the variability among the half-sample estimates of variances comes from the between-strata contributions to these estimates.[6] It was then shown how to eliminate these cross-product terms by choosing a relatively small subset of half-samples in a particular fashion that would yield an unbiased estimator of the true variance of the linear estimator. This variance estimator, in fact, is equal to the value that would be obtained if all possible half-samples were formed to calculate the variance estimate. The half-samples chosen in this manner are said to be orthogonally balanced. This set of half-samples is referred to as an orthogonal set or a balanced set.

The steps involved in composing a balanced half-sample are explained. Assume a sample design of two independently selected units from a stratum. First, the digit 1 or 2 is arbitrarily assigned to each sampled unit from a stratum. Then the number of replicates needed is specified. This number, k, has the restriction that it must be a multiple of 4, in order to obtain an orthogonal set. The value is equal to the number of strata, L, plus either 0, 1, 2, or 3, depending on which integer added to L will yield a number divisible by 4. The next step is to form a $k \times L$ matrix with entries a_{ij} of either plus or minus 1 such that the columns (strata) of the matrix are orthogonal to one another. This means

$$\sum_{\alpha=1}^{k} a_{\alpha j} a_{\alpha j'} = 0 \quad \text{where } j \neq j'$$

Plackett and Burman[13] describe the construction of such a matrix. Finally, the αth half-sample is made by including the unit numbered 1 if $a_{ij} = +1$ and the unit numbered 2 if $a_{ij} = -1$. The names "balanced pseudoreplication" and "balanced repeated replication" are given to the method in which the minimum number of half-samples are needed. (Throughout this text, the term "replication method" means the full orthogonal balanced version unless otherwise specified.)

Also considered in the report[6] was the situation in which the number of strata was so large that the number of half-samples necessary for full orthogonal balancing would be too numerous to be feasible because of computer costs. In this case, a method is suggested whereby the half-samples are partially balanced. To achieve partial balance, the L strata are divided into L/t groups, where t is an arbitrary integer that divides L evenly. The steps given for full orthogonal balancing are carried out for one group of strata, which results in the elimination of the between-strata contributions for these strata. The matrix that is formed is applied to each of the remaining groups. Thus, within orthogonal sets there is no between-strata contribution, and the cross-product terms partially disappear. However, the other cross-products do not drop out. In the case of a stratified random sample and an estimated mean, this subset of half-samples has a lower estimated relative variance than a set of k randomly selected half-samples.

Lee[14] has recently published theoretical and empirical work on a method of selecting a partially balanced set of half-samples. He also considered the problem of the best possible value of t. There will be a larger variance when the replicates are only partially balanced than when they are fully balanced. However, this loss can be reduced when the semiascending order arrangement (SAOA) of selecting half-samples is employed. To accomplish an SAOA, the strata are first arranged in ascending order of the magnitudes of their within-stratum weighted population variance. After this is done, the last $L/2$ or $(L - 1)/2$ (if L is odd) strata are reversed. This means that these strata are put in descending order of magnitude of their within-stratum variance, and the first L/t strata form the first group and so on.

Another factor in reduction is to choose a small t value, but the smaller the value of t, the more half-samples are needed. Therefore, when the investigator decides on the value of t, he must weigh the loss in precision against the increase in computer costs due to a larger number of half-samples.

The behavior of the replication method when a ratio estimator is used was also examined.[15] The findings were primarily based on empirical data from the first cycle of the HES, a sample of approximately 6,600 adults. Body measurements were analyzed from a subsample of approximately 3,000 U.S. adult males.[16] Sixteen regression equations with age, weight, and height as the independent variables and each of the other 16 anthropometric measurements in turn as the dependent variable were tabulated.[c] A variance estimate for a regression coefficient can be produced by the replication method. Using the data in a replicate, a regression equation identical to the one computed for the parent sample can be calculated. Then an estimate of variance is obtained from the average of the deviations of the half-sample regression coefficient estimates from the parent sample estimate. Using a balanced set of 28 replicates, estimates of variances were computed for two types of estimates: ratio estimates of population means for the physical body measurements, and multiple regression and correlation coefficients. According to the study, the set of balanced half-samples gave the same results as would have been obtained by employing the full set of replicates. Also developed was an expression for the relationship between the replication variance of the estimator of the population means and the replication variance of the average of the k half-sample estimated population means. Because the results from a sample provide values for the quantities in the expression, an inference as to whether adequate variance estimates are being obtained can be made. It was concluded that these data showed that the replication variance estimates were sufficient.

Simmons and Baird,[12] following a suggestion by Kish at the Survey Research Center, University of Michigan, investigated whether an approximating variance formula requiring less computer time could be used in place of the exact equation. In the replication method, the data for a half-sample are subjected to the full estimation procedure in exactly the same manner as the data from the parent sample. This operation can use extensive computer time, especially when steps such as ratio adjustment by a concomitant variable and poststratification must be carried out in each half-sample. However, by weighting the individual sample observations more simply than the usual way designated by the sample design, the correct replication variances can be approximated. The goal of the study was to weight the data in three ways to observe whether the approximations could be substituted for the correct variance estimate.

[c]These calculations were provided by the Survey Research Center, University of Michigan, in accordance with NCHS specifications.

The data for the research were the same physical measurements as those used elsewhere.[15]

Operationally, when making the parent sample ratio estimates, a ratio adjustment factor due to the use of a concomitant variable is computed for each of six residence-region classes. This was done only on the data collected in the first cycle of the HES. A poststratification factor for each of 60 age-sex-color classes is also calculated. Both adjustments are in the form of multipliers applied to the weight (reciprocal of the selection probability) for each sampled person. In the replication method, the replicate estimate of the population parameter undergoes the ordinary estimating process. Hence, the multiplication factors of these adjustments should be recalculated in each half-sample to bring the distribution of the sample into as close agreement as possible with the distribution of the universe. By using the two multipliers from the parent sample instead of recalculating them in each half-sample, Simmons and Baird hoped that an adequate approximation involving less computer time could be made. The weighting schemes were to assign each observation (1) a weight of 1 as if the data were collected in a simple random sample, (2) the weight it received when the parent sample estimate was constructed, and (3) the appropriate weight for the specified half-sample containing the observation. A brief summary of the conclusions is offered. The estimates of variance produced from the unweighted data (scheme 1) were not acceptable as approximations to the true variance. The ratios of the variance estimates using weighting scheme 2 to the variance estimates using weighting scheme 1 were all greater than unity, indicating that the simple random sample variance estimates are underestimates. The variability of the ratios for the different statistics was wide. Therefore, the application of a constant adjustment factor to simple random sample variances is not feasible. Simmons and Baird then compared the variance estimates resulting from weighting schemes 2 and 3. The analyses consisted of computing the ratio of replicate variances using the parent sample adjustment factors to variances using unique adjustment factors. These ratios fluctuated around unity. The ratios for the ratio means of the physical body measurements and for the simple correlation coefficients were slightly greater than unity, but the ratios for the partial and multiple correlation coefficients were just below unity. The variability of these ratios was not negligible. Because of this, the investigators concluded that analysis on similar variances for more statistics was required before making a decision about the use of the simple weighting procedure (scheme 2) in place of scheme 3.

The SRC has produced variance estimates by the replication technique since 1957. Kish and Frankel[17,18] presented the empirical evidence on the reliability of the method from SRC's projects and from their analysis of the NCHS body measurement data for the HES. In these studies, the replication method was employed to provide standard error estimates for regression coefficients, correlation coefficients, and partial correlation coefficients. The analyses examined the square roots of the ratios of the true standard errors calculated by balanced repeated replications to variances assuming a simple random sample. These ratios are referred to as the design effect (DEFF).

The first empirical evidence presented was the results from 20 linear regression equations calculated on data for a subgroup of the population sampled in three different surveys. The subgroup was identified as husbands and wives aged 35-44 years living together, with the head of the house in the labor force and having a family income of $3,000 or more. There were 1,853 such defined families. Seven predictors used in various combinations one, two, three, or four at a time were the basis of the regression equations. An estimate of variance for each of 60 regression coefficients was made. The mean value of the square roots of the 60 DEFF's was slightly higher than unity, indicating that simple random sample variance estimates underestimate the true variance.

A total of 1,111 people were interviewed twice to gather data on the political party they supported and their political attitudes. Then, using the political party voted for as the dependent variable and four different attitude scales as predictors, a regression equation was constructed. The mean value of the square roots of these four DEFF's was higher than unity, but lower than the value in the first study. In addition, the mean of the square roots of the DEFF's for the means, the mean value for correlation coefficients, and the mean value for partial correlation coefficients all fluctuated near unity.

Data collected in 3,990 interviews conducted in three national household samples were employed to form six dummy variable regression equations. The predictor variables in this study were education, type of occupation, family income, family reserves, race, and relationship with relatives and friends. For the square roots of 64 DEFF's, the mean value was again above unity.

Kish and Frankel,[17] under a contract with NCHS, estimated in the 16 linear regression equations based on the physiological measurements of 3,091 males in the HES the ratio of the replication variance to simple random sample variance. The mean value here for the regression coefficients was above 1 and higher than the previous empirical values. The investigators suggested that this higher mean value resulted from the large clusters used in the HES.

The last set of empirical results examined the variance estimates for statistics computed by the technique of multiple classification analyses. This technique is a multivariate analysis that examines the relationship of each predictor to a dependent variable that must be an interval scale or a dichotomy, with the other predictors held constant. The model is additive; and no assumptions of equal cell frequencies, linearity of regression, and orthogonality are required. For each predictor, there are two estimated statistics, eta and beta coefficients. The squared eta coefficient is the proportion of variance of the dependent variable explained by unadjusted deviations (cell means minus overall mean), and the squared beta is the proportion of variance of the dependent variable explained by a given predictor, holding the other predictors constant. Data for a sample of 2,214 family heads were used in the multiple classification analyses equation to fit a receptivity index to education of head of household, age of head, total family income, social participation, achievement orientation, sex, and marital status. Because of the cost involved, the replication variance estimates were computed from 12 partially balanced replicates and the simple random sample variance estimates from 12 repetitions based on simple random splits of the data. For the six eta coefficients, the mean DEFF's were greater than 1, as were the mean DEFF's for the corresponding betas.

By means of a Monte Carlo approach, Levy[19] compared the performance of the balanced replication and jackknife methods for the ratio estimator when the sample design was stratified, with two PSU's selected from each stratum. Within each of 16 strata, a synthetic normal population with specified mean and variance was generated for two variables X_h and Y_h. The parameter to be estimated from the samples was

$$R = \frac{\sum_h X_h}{\sum_h Y_h},$$

a ratio estimator. The sampling and estimation procedure was to draw from each stratum two independent estimates of X_h and two independent estimates of Y_h. These estimates were fashioned into a combined ratio estimate, $\hat{R}$. From a balanced set of 16 half-samples, an estimate of the variance of $\hat{R}$ was computed. Another estimate of this variance was calculated by the jackknife method.

In the experiment, the parameters R_h were allowed to vary in succession according to the ratios 1.01, 1.05, and 1.10 for one value of the correlation between X and Y and one value of the relative variance of the population mean X. The correlation value used equaled 0.9, and the relative variance was set equal to 0.01.

For analytical purposes, the estimated variances, biases, and mean square errors for both the replication and the jackknife method were expressed relative to the sampling variance of the 1,200 estimates of $\hat{R}$. The relative variance equaled the sampling variance of the variance estimates for a method divided by the square of the sampling variance of $\hat{R}$. The relative bias squared equaled the square of the mean of the variance estimates for a method minus the sampling variance of $\hat{R}$ divided by the square of the sampling variance of $\hat{R}$. The relative mean square equaled the sum of the relative variance plus the relative bias squared.

The relative mean square errors for both methods increased slightly as the varying ratio increased from 1.01 to 1.10. The jackknife relative mean square error estimates were lower than the estimates produced by replication for all three cases. The jackknife estimates of relative variance were also lower than replication estimates for all the varying ratios. When the R'_h's varied by the ratios 1.05 and 1.10, the replication method had the lower relative bias. These results indicate that further such studies using different parameter values are needed.

Several of the design and estimation components included in complex sample surveys are unequal sampling fractions, clustering, stratification, first-stage ratio adjustment, nonresponse adjustment, and poststratification. In an empirical investigation, Simmons and Bean[20] measured the effect these components, alone and in combination, had on replication variance. Data for the four health conditions—hypertensive heart disease, myocardial infarction, angina pectoris, and syphilis—from the HES were used. Using six estimation procedures, six estimates of the proportion of the total population or proportion of a subclass of the population having a specified health condition were calculated. To implement these procedures, each examined person received a set of six weights. The weights assigned were (1) a weight of 1, as if the data were collected in a simple random sample; (2) a weight of 10,000; (3) the appropriate weight for the HES sample design; (4) the weight in 3, multiplied by a nonresponse adjustment; (5) the weight in 4, multiplied by a first-stage ratio adjustment; and (6) the weight in 5, multiplied by a poststratification factor. A simple random-sampling variance estimate was calculated for the estimate produced when the weight of 1 was used; replication variances were tabulated for the other five estimates.

Ratios of variances and ratios of mean square errors formed the basis of comparison. For the ratios of variances, the numerator was the variance of one of the estimates, and the denominator was the variance of another estimate. For the ratios of variances, the numerator was the variance estimate produced by one method, and the denominator was the variance estimate produced by one of the other methods. To take bias into account, the ratio of the mean square error of one estimate to the mean square error of another was examined. The estimate produced by the full design and estimation procedure (the last weighting scheme) was substituted for the true value. Thus, an estimate of bias equaled the difference between this estimate and the estimate obtained for the method in comparison.

The results of the comparison based on ratio of variances and the general conclusions are presented. The use of unequal sampling fractions alone showed a 27-percent increase in replication variance over simple random sample variance. The combined effect of unequal weighting, clustering, and stratification produced a 137-percent increase in replication variance over a simple random sample model. However, the variance estimate for the estimate produced by the full design and estimation procedure was only about double the simple random variance estimate. Thus, the estimation steps—nonresponse, first-stage ratio adjustment, and poststratification—improved the precision of the estimates. Simmons and Bean concluded that none of the design and estimation components investigated introduced large biases in the estimates and that these components have substantial impacts on variance estimates.

The strong point of the replication variance estimator method is that measurement errors, design components, and estimation procedures are all accounted for. The technique can be employed to yield variance estimates for any statistic and, thus, is flexible. It can also be adapted to a design that selects more than two PSU's from each stratum.

Linearization Method

The other general method for estimating variances was developed by Keyfitz,[21] who outlined a method for estimating the variance of a poststratified estimator. This estimator is derived from the fact that the variance of a sum of two independent estimates of a parameter equals the expected value of the square of the difference between them. This is

$$\text{VAR}(X_1' + X_2') = E(X_1' - X_2')^2$$

where X'_1 and X'_2 are estimates of the parameter X made from two independent random samples drawn with replacement. The key theorems in developing the method will be sketched without giving any proofs. This expression can be extended to include estimates drawn from another stratum, assuming the selections among the strata are independent.

The next term needed is the covariance between characteristics measured on the same sampled units within a stratum. Assume X'_1 and Y'_1 are estimates from one random sample and X'_2 and Y'_2 are estimates from another random sample. The two samples are drawn independently with replacement and, thus, $\mathrm{Cov}(X'_1, Y'_2) = \mathrm{Cov}(X'_2, Y'_1) = 0$, $E(X'_1) = E(X'_2)$, and $E(Y'_1) = E(Y'_2)$. Then the covariance can be shown to equal

$$\mathrm{Cov}(X'_1 + X'_2),(Y'_1 + Y'_2) = E(X'_1 - X'_2)(Y'_1 - Y'_2)$$

The key in the derivation of the linearization method is the approximation of the relative variance of a ratio estimator using the Taylor series expansion. Defining the relative variance of an estimator as the variance of the estimator divided by the square of the expected value of the estimator, then relative variance (V^2) is

$$V^2\left(\frac{X'}{Y'}\right) = E\left[\frac{X'}{E(X')} - \frac{Y'}{E(Y')}\right]^2$$

$$\simeq V^2(X') + V^2(Y') - 2V(X',Y')$$

where

$\frac{X'}{Y'}$ = the ratio estimate of the population parameter $\frac{X}{Y}$,

$V(X',Y') = \rho(X',Y')V(X')V(Y')$,

$\rho(X',Y')$ = the correlation between X' and Y',

$V(Y')$ = the coefficient of variation of Y', and

(X') = the square root of $V^2(X')$.

A proof of this result is in Hansen, Hurwitz, and Madow.[22]

Assuming again independent random samples drawn with replacement and using the above theorems, substitution, and algebraical manipulations, it may be shown that

$$V^2\left(\frac{X'_1 + X'_2}{Y'_1 + Y'_2}\right) = E\left[\frac{X'_1 - X'_2}{E(X'_1 + X'_2)} - \frac{Y'_1 - Y'_2}{E(Y'_1 + Y'_2)}\right]^2$$

If selections are independent among strata, these results can be generalized to any number of strata.

The last theorem gives the variance of a poststratified estimator. The sample design is two random selections drawn independently and with replacement from each stratum. The poststratified estimator is

$$X'' = \sum_a \frac{\sum_h (X'_{ha1} + X'_{ha2})}{\sum_h (Y'_{ha1} + Y'_{ha2})} Y_a$$

where

X'' = the poststratified estimate of X characteristic,

X'_{hai} = the ith estimate for the ath class of the hth stratum,

Y'_{hai} = the ith estimate of population for the ath class of the hth stratum,

Y_a = precalculated total population in the ath group.

The difficulty in determining the variance of X'' is the lack of independence among the classes. The variables defining the classes are not used to stratify the population. Therefore, the covariance between the classes is not zero. This method evaluates the covariance between the groups separately in each stratum. These values are then summed across the strata. The variance of X'' is

$$\mathrm{VAR}(X'') = E\sum_h \left[\sum_a Y_a \frac{E\sum_h (X'_{ha1} + X'_{ha2})}{E\sum_h (Y'_{ha1} + Y'_{ha2})} \right] \times \left[\frac{X'_{ha1} - X'_{ha2}}{E\sum_h (X'_{ha1} + X'_{ha2})} - \frac{Y'_{ha1} - Y'_{ha2}}{E\sum_h (Y'_{ha1} + Y'_{ha2})} \right]$$

In this discussion, random sampling with replacement has been assumed. If sampling is without replacement, the finite population correction factor can be inserted into the formulas. From these equations, formulations applicable to complex sample designs have been developed. Tepping[23] developed a more general procedure for linear variance estimators using a Taylor series expansion and showed how the method can be employed to provide variance estimates for more complicated statistics such as regression coefficients. Bean presented the theory, proofs, and description of the form of the variance estimator for the ratio estimates produced by the HIS report.[24] Shapiro[25] and Waltman[26] outlined the adaptation of Keyfitz' theorems to the ratio estimation procedures employed in the CPS. In essence, the linearization method fashions a variance estimator from a linear combination of sample totals for each PSU.

Statisticians of the Bureau of the Census were among the first to apply the linearization method. They investigated its use in the CPS's sample design and estimation procedures and constructed a computer program tabulating the values needed for each PSU. Banks and Shapiro[8] presented and discussed the empirical basis the Bureau had acquired over the years from its variance estimates for ratio estimators. They also assembled the data on the replication variance estimates. The methodology used in the analysis was the comparison of design effects for unbiased estimates, ratio estimates, and composite estimates for a variety of labor force variables. The values (true variance as computed by linearization technique) for the unbiased estimates and ratio estimates indicated that, without exception, the ratio estimate lowered the variance. In fact, for the variables with large parameter values, the decrease was substantial. When the comparison was made between the ratio estimates and the composite estimates, the latter reduced variance of some of the labor characteristics, but not all of them.

Data from 1964 on 13 variables, the relative variances as estimated by linearization and replication (not balanced) methods, were displayed in Banks and Shapiro. Both estimates were consistently of similar size, but there was more variability among the monthly replication variance estimates. The authors concluded that the linearization estimates were more reliable.

Linearization is the device by which variance estimates of the ratio-estimated characteristics of the Canadian Labor Force Survey are calculated. Fellegi and Gray[27] explained the application of this method to the survey with its particular sample design and estimation procedure and discussed how the analysts actually make use of the design effect in their analysis. The comparative findings of the design effects were consistent with the data from the U.S. Bureau of the Census study.

The literature on linearization is small compared to that on replication. There are probably two reasons for this. Until Tepping's article,[23] this method was not available for the estimation of variances for statistics other than means, and the method did not generate the interest of the replication scheme. Furthermore, samplers may have used the technique without reporting its general behavior. One feature of linearization is that the variance can be estimated for each stage of sampling, a fact that is important for designing surveys.

An Empirical Investigation of All Three General Variance Estimators

In the articles reviewed, there are few comparisons of the properties and behavior of the estimates produced by the two methods when they are applied to the same data. The only extensive research along these lines has been the work of Frankel.[28] Using repeated sampling from a universe modeled from real sample survey data, the validity of three fundamental assumptions, important for purposes of statistical inference, was assessed. The assumptions were the following:

1. The sample estimate of the population parameter is approximately unbiased.
2. An approximately unbiased estimate of the variance of the estimate is computable from the sample.
3. The distribution of the ratio of the sample estimate minus its expected value, to its estimated standard error is reasonably approximated by the "Student" t within symmetric limits.[d]

Frankel regrouped the CPS's March 1967 data on noninstitutional families and individuals into PSU's consisting of three segments of approximately five to six households each. These segments were the basic sampling units of CPS. The PSU's were stratified into 6, 12, and 30 strata, and two PSU's were selected independently from each stratum for each design. For design I (6 strata) and design II (12 strata), 300 samples were drawn; for design III (30 strata), 200 samples were chosen. Thus, the sample design consisted of one stage of sampling, stratification, and single-stage clustering.

The estimation procedure followed in each sample was the fitting of two linear regression equations. The total income of household head was related to the number of persons in the household under 18, the number of persons in the household, and the sex of the household head. For the second equation, the total income of the household was the independent variable, and the dependent variables were number of persons in household in labor force, age of household head, and years of school completed by household head. Estimates were made of simple means, differences of means, correlation coefficients, regression coefficients, partial correlation coefficients, and multiple correlation coefficients. For each of these estimates, the variance was calculated in nine different ways. The nine methods included all three of the general variance estimators: Taylor expansion (linearization), balanced repeated replication, and jackknife repeated replication.[e] A description of the specific jackknife formulation used can be found in Frankel's[28] work.

Frankel compared the behavior of the estimates of means, difference of means, regression coefficients, and simple, partial, and multiple correlations on the basis of relative bias. Relative bias is defined to be the difference between the estimate of the parameter and the parameter divided by the parameter. The relative biases of the mean and differences of means were all less than 1 percent, except for one difference of means that had a relative bias of 3 percent. The regression coefficients had higher relative biases than did the means. In design I, seven out of the eight regression coefficients showed relative biases of less than 4 percent; in design II, only five out of the eight were less than 4 percent; in design III, the relative biases for seven of the regression coefficients were less than 4 percent. The correlation coefficients (simple, multiple, and partial) displayed the highest relative biases. Only for design III did the mean of the relative biases for simple correlations drop below 5 percent.

In analysis of assumption 2, the nine variance estimators were also considered to be mean square error estimators, and their performances in that capacity were studied. The relative bias was the main measure of evaluation. For means and differences of means, none of the nine had minimum relative bias across all the designs; for simple correlations, one of the replication methods emerged with lowest relative bias. However, for the estimates of regression coefficients, a jackknife technique displayed minimum relative bias for two designs and linearization behaved best for the other design. In all three designs, the replication estimate, involving the square of the differences between half-sample estimate and complement estimate, showed the lowest relative bias for partial correlation coefficients.

The results provided evidence for the validity of assumption 2, but the data did not definitely

[d]Frankel,[28] p. 2.

[e]There were nine methods because each of the last two techniques provide four different possible schemes of variance calculations (the four produced by replication are outlined in the Review of Literature section).

resolve the issue of which variance scheme produced estimates closest to the true variance. The conjecture that the estimates produced by the three general methods are really estimates of mean square errors rather than variances was studied using another relative bias expression. For designs I and II, variance estimates for means and differences were relatively closer to the mean square error parameter than the variance parameter. More research is required before a definite answer can be given as to whether the estimates produced by the general methods are mean square error estimates rather than variance estimates.

Probably the most important conclusion reached concerned assumption 3. The ratios of the sample estimate minus its expected value divided by each of its nine estimates of standard errors were computed. The evaluation of the ratios was based on the proportion of times the ratios fell within symmetric intervals around the origin. The intervals chosen were normal values. One-sided intervals were also computed, and examination of the proportions within these intervals revealed instances of asymmetry. Frankel concluded that these findings were partly real and partly the result of selecting only several hundred samples rather than several thousand samples. Because of this, he decided not to compare the proportions with values in the *t* table. To reduce variability of the ratios, proportions were averaged for similar statistics within a design. These proportions were compared to the value in the *t* table with degrees of freedom equal to the number of strata. The findings indicated that the *t* distribution within symmetric intervals could be used to make inferences. Another vital observation was that ratios based on the replication scheme were more likely to be in agreement with the proportions of that distribution than the proportions based on the estimated standard errors.

The Frankel investigation has contributed significantly to knowledge concerning the properties of the variance estimates produced by the three general variance estimators. The validity of the assumptions investigated needs to be confirmed for different sample designs, for different variables, and for different estimators.

METHOD

Description of Population

Morbidity data collected by the HIS in 1969 on 131,575 civilian, noninstitutional U.S. individuals constitute the universe for this investigation. The HIS[29] is a highly stratified, multistage annual probability sample that provides estimates of disease incidence and prevalence and health characteristics.

The same basic sampling design has been used since the survey began in 1957. The first-stage unit, PSU, is a single county or several contiguous counties. The entire geographical territory of the United States is classified into 1,900 such units. The next step is the stratification of the 1,900 units into 357 strata, many of which consist of only one PSU. Geographically, the PSU's are divided into subunits (each containing an expected six households) called segments. Sampling for the survey is done in two stages: one PSU is selected with probability proportional to population size from each of the 357 strata. The PSU's comprising an entire stratum are included with probability 1 and are referred to as self-representing PSU's (SR PSU's). The other PSU's are called non-self-representing PSU's (NSR PSU's). Within each selected PSU, segments are interviewed. An interview schedule on every individual residing in the household is completed. Thus, the design includes the effects of stratification and two stages of clustering, providing the model for this study.

The features of the sample design employed in this empirical investigation were stratifying the PSU's, sampling of two stages, and clustering at two levels. The steps involved in implementing the design were to sample the first-stage units with probability proportional to population, the largest units entering the sample with probability 1, and then to subsample within the selected first-stage units. However, before any of the details of the sample design—e.g., the decision on the sampling fraction to be used—could be determined, the HIS data had to be regrouped.

The 357 selected PSU's, each representing a stratum in HIS, were taken as the total population. An examination of the 1969 data for the original Health Interview Survey PSU's revealed a mean of only 18 segments per PSU, excluding from the average eight PSU's having over 100 segments. This average segment size was too small to support a two-stage design. Therefore, the 357 units were regrouped to form 149 PSU's for the universe. Paralleling the HIS design, the 149 PSU's were classified into 19 strata, eight consisting of only one PSU, grouped primarily by geography and population size. The original segments were retained, except the real outliers (segments with three or fewer persons) were combined with other segments to reduce variability of segment size within PSU.

The HIS observations were for persons, but since in this investigation everyone within a selected segment was to be in the sample (eliminating the problem of adjusting for nonresponse), the person records were combined to form a single segment record. To recapitulate the construction of the population, the 357 PSU's in the HIS sample were reorganized into 149 units, the units were categorized into 19 strata, the segments were edited for outliers, and the data collected on persons in each segment were combined into a single segment record.

There are several reasons for using these data as the population. The data had already been collected, edited, and made available on magnetic tape. However, even after the completion of these operations, further extensive manipulations were required to put the data into the form required to perform the necessary calculations. One purpose of the project was to measure the effects of stratification and clustering on the variance estimators studied. The data reflect the real homogeneity or clustering that exists in human populations. The estimation component, poststratification, which was to be investigated to ascertain whether the variance estimators were sensitive to estimation procedures, could easily be performed on these data. Another basis for their use was the variables collected in the survey. Evidence of the behavior of the balanced half-sample replication and the linearization method is needed for a variety of variables in order to determine if observed differences between the variance estimators produced by the two methods are caused by the nature of the variables themselves or by the methods. The variables available from HIS data are primarily morbidity information based on an individual's own perception of his health.

Sample Design

As stated earlier, the population was divided into 19 strata, 8 self-representing and 11 non-self-representing. Characteristics of the universe are given in table 3. First-stage sampling consisted of the selection of the eight SR PSU's and the selection of two PSU's from each of the 11 NSR stratum.

The number of PSU's in each NSR stratum was small; thus, the finite population correction factor (fpc factor) would have a considerable impact on the variances. The usual procedure would have been to sample without replacement and then include the fpc factor in the formulas for the variance estimator, but unfortunately, the theory for incorporating fpc factor in these calculations has not been developed. The solution was to sample with replacement. One desirable feature of a design is for the sample to

Table 3. Characteristics of the population

Stratum	Number of primary sampling units	Number of segments	Population size
1	1	323	5,219
2	1	324	5,295
3	1	293	5,097
4	1	257	4,612
5	1	270	4,649
6	1	191	3,338
7	1	339	5,519
8	1	271	4,517
9	14	593	10,238
10	15	592	10,573
11	13	526	8,815
12	12	482	7,638
13	13	461	7,332
14	11	449	7,809
15	12	481	7,785
16	11	446	7,854
17	13	455	7,679
18	13	502	8,997
19	14	513	8,610
Total	149	7,768	131,575

be self-weighting, which means that each element has the same probability of being selected. This feature can be included when the PSU is sampled with probability proportional to size. The first stage of sampling included the selection of two PSU's independently from the 11 NSR strata with probability proportional to size and with replacement.

To determine whether sample size plays a part in the observed differences in the variance estimates produced by the methods, three sample designs were used. The number of PSU's sampled was constant, so the subsampling rate was varied to achieve different overall sample sizes. To accomplish this, a uniform sampling fraction f, equal to the product of the PSU selection probability of selecting the PSU times the subsample fraction within the PSU, was used in each stratum. The formula is

$$f = f_1 \times f_2$$

where

f = uniform sampling fraction,

f_1 = the sampling rate of first-stage units, and

f_2 = the sampling rate of second-stage units.

In design I, f equaled 1/75; in design II, 2/75; and in design III, 2/30. The fraction f_2 equaled f/f_1. Thus, by setting f and knowing f_1, the f_2 could be computed. Because two PSU's were chosen in the NSR strata, the values 1/150, 1/75, and 1/30 were used in the computations. The expected yields from the two fractions of designs I and III were 4 segments and 20 segments, which approach the two possible extremes in sampling: simple random and ultimate cluster. The optimum design probably falls between the two rates used. The second stage of sampling was random subsampling with replacement of the segments at the rate f_2 from the chosen PSU's.

To clarify the sampling plan, consider design III, which had an overall sampling fraction of 2 in 30. First, the eight SR strata entered the sample with probability 1. Within each, segments were randomly selected at the rate f_2 = (2/30)/1 = 2/30. Thus, for stratum 1, approximately 22 segments were drawn. Next, two PSU's were chosen with probability proportional to population from each of the remaining 11 strata. Within a selected PSU, the segments were subsampled with the rate $f_2 = (1/30)/f_1$.

For each design, 900 samples were independently drawn. In drawing these samples, not only was the sampling distribution of the variance estimators being estimated, but also being estimated was the sampling distribution of the ratio estimators $\hat{R}_i$, i=1, · · · , 900. In repeated sampling, this ratio estimator will settle down to its expected value, and, thus, the simple sample variance

$$\sum_{i=1}^{900} (\hat{R}_i - \bar{\hat{R}})^2 / 899$$

will become fairly stable. The value 900 was thought to be an adequate number to achieve stability.

Variables

Five variables were selected from the HIS to represent a variety of basic distributional forms. These are as follows:

Variable[f]

Family income
Number of restricted activity days in past 2 weeks
Number of physical and dental visits in past 12 months
Number of days spent in short-stay hospitals in past 12 months
Whether or not the person has seen a physician in last 12 months

Quantity estimated

Average income per person
Average number of restricted activity days per person per year
Average number of visits per person per year
Average number of short-stay hospital days per person per year
Proportion of population seeing a physician in last 12 months

Two criteria for selection of the variables were that (1) they could be obtained from the person-record HIS tapes and (2) they would have a variety of distributions. Ideally, what is

[f]Each sample person's record contained a measure for each of the five variables; therefore, to obtain a segment record, these measures were summed across all persons in the segment.

desired is for the computed variance estimates to reflect the differences of the variance estimators. Thus, the possibility that the variables themselves cause the observed differences between the variance estimates is investigated.

It is important to remember in considering the variables measured in this study that the ultimate sampling unit is the segment and that although several variables are converted to a per-person basis, no direct assessment of within-segment variability is available. Thus, distributions of variables in the study directly reflect segment-to-segment variation.

The income variable is a family income figure. Thus, each person within a family is assigned the same value. Income, then, was a highly clustered, continuous variable. However, in the HIS, income was reported only in group intervals, with the highest class being open ended. To convert from grouped data, midpoints of each interval, except for the open one, were the values given to individuals. Everyone having a family income falling into the open interval was assigned the lowest income of the class.

The proportion of persons in the population visiting a physician in the past 12 months was selected as an example of a variable distributed on the unit interval. The variable, number of restricted activity days, produces an incidence statistic based on the respondent's recall of the past 2 weeks. The respondent's recall for the remaining two variables is 12 months. The variable, number of days spent in short-stay hospitals in the past 12 months, has a J-shaped distribution. Most individuals do not spend any days in the hospital, but there are a few people whose hospital experience lasts for months. These two extremes account for the J-shape.

These variables provide an example from each type and each range class of statistics estimated in the HIS. The statistics produced by the HIS are classified by the length of the recall period and by the usual value of the measure of a health characteristic for an individual. The classes are (1) narrow range—the measure is usually 0 or 1 and occasionally 2, e.g., the number of days spent in short-stay hospitals; (2) medium range—the typical value for an individual is from 0 to 5, e.g., the number of physician and dental visits; and (3) wide range—the measure for an individual is usually greater than 5, e.g., the number of restricted activity days.

ESTIMATORS AND VARIANCE ESTIMATORS

Estimators

Two estimating equations, each producing from the sample a ratio poststratified estimate of the population ratio parameter, were employed with the level at which poststratification was done as the differentiating characteristic. Each equation consisted of three basic operations: inflation by the reciprocal of the probability of selecting second-stage units, inflation by the reciprocal of the probability of selecting first-stage units, and poststratification.

The purposes of stratification are to improve precision of the overall population estimates and to insure that subgroups of the population (domains of study) are in the sample in the same proportions as they are in the universe. Usually, the universe cannot be stratified into these subgroups before sampling. However, if the distribution of the subgroups in the universe is known, the sample results can be adjusted (poststratified) so that these domains of study are appropriately represented and the accuracy of the estimate increased.

The subgroups in this study were the population in 24 ethnic-sex-age classes (white and nonwhite, male and female, ages 0-4, 5-14, 15-24, 25-44, 45-64, and 65+). This spread resulted in small frequencies in most of the nonwhite cells, another reason for poststratifying. Typically, the distribution of the subgroups is known only for the complete population, and adjustments must be made at this level. In this situation, however, the universe was totally specified and poststratification could be carried out either after combining the stratum estimates for an ethnic-sex-age class or earlier. Poststratification was applied at two different levels: universe and region.

Universe level meant that the sample estimates for a class were combined across all the strata and then adjusted to the total population in that group. Region-level poststratification refers to the adjustment of sample estimate for the PSU's within a region to the regional distribution of classes. The PSU's belonged to one of two regions: the combined East and North Central regions or the combined South and West regions.

The first estimator of a statistic (see appendix I for the formula for $\hat{R}_2$) is

$\hat{R}_1$ —poststratification at universe level

$$\hat{R}_1 = \frac{x_1''}{y} = \frac{\sum_{a=1}^{24} x_a' \frac{y_a}{y_a'}}{y}$$

where

$\hat{R}_1$ = the final poststratified ratio estimate of the population parameter R,

x_1'' = the poststratified estimate of health characteristic x,

y = the total population size,

x_a' = the inflated estimate combined across all PSU's of health characteristic x for ethnic-sex-age class a,

y_a = the known population in the ethnic-sex-age class a, and

y_a' = the inflated estimate across all PSU's of the population in class a.

Variance Estimators

The main objective of the study is to examine the behavior of the variance estimators produced by two general variance estimation methods, replication and linearization. Ideally, the variance of each estimator produced by one of the two estimating equations would be estimated using the correct application of both variance formulas and each of the approximate formulas, but due to computer costs and volume of numbers, this was not feasible. For estimator 1, variance estimates were computed by the correct replication and linearization methods and two approximate schemes, while for estimator 2, only the correct replication method was used. (See appendix II for the formulas for the approximate variance estimates and for the correct replication variance estimator for $\hat{R}_2$.)

An assumption of both general methods is that two PSU's are chosen independently from each stratum. The sample design for this study with SR PSU's violates the assumption giving rise to the problem of how to treat these PSU's. The solution was to divide randomly the sampled segments of each SR PSU into two groups and to assume that these two groups, pseudo-PSU's, were two independent selections from the same stratum.

Variance Estimators for Estimator 1

The variance for the estimate $\hat{R}_1$ was estimated in 10 different ways. Three replication methods, each yielding three estimators, and one linearization method were used.

The process for computing the correct replication variance estimator is as follows: Assume that from each stratum one of the two sampled primary units is selected. Nineteen PSU's form a half-sample. The data from these 19 PSU's are inflated by the reciprocal of the probability of selection to their stratum level. These estimates for a particular ethnic-sex-age class are combined across all the strata. Then, within the classes, the estimates are adjusted by poststratification factors. These factors are calculated, using the population estimates for the class produced by only the PSU's in the half-sample, on the universe level, in the same manner used to calculate the factors for the parent sample. The 19 PSU's in the complement replicate undergo an identical process to yield a similar estimate of the parameter. This is repeated for 20 such half-samples. The three variance estimates produced for each sample estimate obtained from estimating $\hat{R}_1$ are

$$\mathrm{VAR}(R1H) = \frac{1}{20} \sum_{\alpha=1}^{20} (R_{1,\alpha}' - \hat{R}_1)^2$$

$$\mathrm{VAR}(R1C) = \frac{1}{20} \sum_{\alpha=1}^{20} (R_{1,\alpha}^* - \hat{R}_1)^2$$

and

$$\mathrm{VAR}(R1S) = \frac{1}{2}\left[\mathrm{VAR}(R1H) + \mathrm{VAR}(R1C)\right]$$

where

$\hat{R}_1$ = the parent sample ratio estimate,

$R'_{1,\alpha}$ = the αth half-sample ratio estimate of the population parameter R by an identical estimation equation to the one that yielded $\hat{R}_1$, and

$R^*_{1,\alpha}$ = the αth complement half-sample ratio estimate.

The 20 half-samples are composed from the orthogonal pattern in table 4. This design was constructed by following the method described in Plackett and Burman.[13] The pattern is determined by the rotation of the first 19 entries of the first column and by specifying the 20th row to always be minus.

The linearization formula employed in this study was derived from the theorems given by Keyfitz.[21] The form of the method used here can be described as follows: Remembering that $R_1 = x''_1/y$, the formula for the aggregate x''_1 is

$$x''_1 = \sum_{a=1}^{24} x'_a \frac{y_a}{y'_a}$$

This equation can be rewritten as linear combinations of the simple inflated estimates of the PSU's from each stratum. Expressed this way, the equation becomes

$$x''_1 = \sum_{a=1}^{24} \frac{\sum_{h=1}^{19} (x'_{ha1} + x'_{ha2})}{\sum_{h=1}^{19} (y'_{ha1} + y'_{ha2})} y_a$$

Table 4. Orthogonal pattern for 20 replicates[1]

Replicate	Stratum																		
						SR PSU's									NSR PSU's				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−
2	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+
3	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+
4	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−
5	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−
6	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+
7	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+
8	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+
9	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+
10	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−
11	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+
12	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−
13	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+
14	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−
15	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−
16	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−
17	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−
18	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+
19	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+
20	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

Note.—SR PSU indicates self-representing primary sampling unit; NSR PSU indicates non-self-representing primary sampling unit.

[1] "+" denotes +1 and "−" denotes −1.

where

x'_{hai} = the inflated estimate of health characteristic x for class a of the ith PSU in the hth stratum and

y'_{hai} = the inflated population estimate for class a of the ith PSU in the hth stratum.

From this expression, the linearization variance estimator of $\hat{R}_1$ is derived. The formula is

$$\mathrm{VAR}(L1) = \frac{\mathrm{VAR}(x''_1)}{y^2}$$

$$= \frac{\sum_{h=1}^{19}\left[(x'_{h1} - x'_{h2}) - \sum_{a=1}^{24} \frac{x''_{1,a}}{y_a}(y'_{ha1} - y'_{ha2})\right]^2}{y^2}$$

where

x'_{hi} = the inflated estimate for the ith PSU of the hth stratum,

$x''_{1,a}$ = the final poststratified ratio estimate of the health characteristic x for the ath ethnic-sex-age class.

For more detailed formulas and discussion, see Bean.[30]

RESULTS

Introduction

The main objective of the investigation was to determine if either or both of the general variance estimator methods, replication and linearization, yields acceptable variance estimates for a complex ratio estimator employed in multistage probability sample surveys. The properties considered critical for acceptability are the extent of the bias of the variance estimates, the degree of their variability, and the amount of their mean square error. Another very important point is whether inferences can be made when variance estimates are used. The other result presented is the effect of the estimation technique poststratification on variances. The bias of the two ratio estimators employed and the substitution of approximate variance estimates are discussed in appendixes I and II. Bean[30] presents a more in-depth report on other properties of the ratio estimators, other approximate variance estimate methods, and the results, using several different procedures for estimating the variance component from SR PSU's.

Comparison of Replication and Linearization

The primary purpose of this investigation was to compare replication and linearization methods of variance estimation for a sample design that includes stratification, two stages of clustering, and poststratified ratio estimation. The comparisons are in terms of bias, variance, and mean square error.

Only for the sample estimator $\hat{R}_1$ were both the appropriate replication and linearization variance estimates computed. "Appropriate" here refers to the best application of the theory of the methods to the particular sample design and estimation process employed. For the replication method, this means that the data for the PSU's in a half-sample were (1) inflated to the PSU level, (2) inflated next to the stratum level, and (3) poststratified by adjustment factors computed from the half-sample itself. Therefore, the half-sample estimate of the population parameter contained all the elements of the sample design and estimation procedure. For the best application of the linearization theory, the basic equations by Keyfitz[21] were interpreted to include inflation to PSU and stratum levels and to include an adjustment for poststratification. For each estimate of $\hat{R}_1$, one linearization and three replication variance estimates were calculated. The replication estimates VAR($R1H$), VAR($R1C$), and VAR($R1S$) all have the same expected value. Because VAR($R1S$) is an average of the other two, its precision is expected to be greater. Since the estimator VAR($R1H$) is the one most often used by statisticians, the discussion will deal with it and with VAR($R1S$).

One important question is whether the methods yield biased statistics. To compute true bias, the real variance of $\hat{R}_1$ for this sample design is needed. Unfortunately, this value was not known, but with 900 samples, the distribution of $\hat{R}_1$ should become quite stable; and the sample variance of 900 $\hat{R}_1$ estimates should be a good criterion. The sample variance of $\hat{R}_1$ is

$$s^2_{\hat{R}_1} = \frac{1}{899} \sum_{j=1}^{900} (\hat{R}_{1j} - \bar{\hat{R}}_1)^2$$

where

$$\bar{\hat{R}}_1 = \frac{1}{900} \sum_{j=1}^{900} \hat{R}_{1j}$$

By assuming $s^2_{\hat{R}_1}$ to be the true variance, measures of bias can be calculated. These are

$$\text{bias } [\text{VAR}(R1H)] = \overline{\text{VAR}(R1H)} - s^2_{\hat{R}_1}$$

where

$$\overline{\text{VAR}(R1H)} = \frac{1}{900} \sum_{j=1}^{900} [\text{VAR}(R1H)]_j,$$

$$\text{bias } [\text{VAR}(R1S)] = \overline{\text{VAR}(R1S)} - s^2_{\hat{R}_1}$$

where

$$\overline{\text{VAR}(R1S)} = \frac{1}{900} \sum_{j=1}^{900} [\text{VAR}(R1S)]_j,$$

and

$$\text{bias } [\text{VAR}(L1)] = \overline{\text{VAR}(L1)} - s^2_{\hat{R}_1}$$

where

$$\text{VAR}(L1) = \frac{1}{900} \sum_{j=1}^{900} [\text{VAR}(L1)]_j$$

Inspection of table 5 reveals that the bias of the $R1H$ estimator decreases across the three designs for four of the five variables. Only for the variable proportion seeing a physician does the bias, which goes from -6.083×10^{-6} in design I to 1.614×10^{-5} in design II and then down to 4.206×10^{-6} in design III, not follow the pattern. For physician visits and hosp'tal days, the bias is always positive, while for family income and restricted activity days, it switches from positive to negative.

Table 5. The bias of the $R1H$, $R1S$, and $L1$ variance estimators of the sample estimator $\hat{R}_1$ of the population parameter R

Variance estimator	Variable				
	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R1H$)	3.857×10^{3}	2.736×10^{-1}	2.751×10^{-3}	3.750×10^{-3}	-6.083×10^{-6}
VAR($R1S$)	4.103×10^{3}	2.783×10^{-1}	2.558×10^{-3}	3.765×10^{-3}	-7.095×10^{-6}
VAR($L1$)	-3.191×10^{3}	-1.085×10^{-1}	-4.050×10^{-3}	-2.242×10^{-3}	-4.944×10^{-5}
	Design II				
VAR($R1H$)	1.489×10^{3}	-4.337×10^{-2}	2.693×10^{-3}	1.929×10^{-3}	1.614×10^{-5}
VAR($R1S$)	1.640×10^{3}	-4.092×10^{-2}	2.688×10^{-3}	1.785×10^{-3}	1.581×10^{-5}
VAR($L1$)	-3.211×10^{2}	-1.598×10^{-1}	1.022×10^{-3}	-6.848×10^{-5}	6.945×10^{-6}
	Design III				
VAR($R1H$)	-8.772×10^{2}	-2.934×10^{-2}	1.073×10^{-3}	4.490×10^{-5}	4.206×10^{-6}
VAR($R1S$)	-8.909×10^{2}	-2.900×10^{-2}	1.042×10^{-3}	3.684×10^{-5}	4.069×10^{-6}
VAR($L1$)	-1.271×10^{3}	-5.818×10^{-2}	7.121×10^{-4}	-3.094×10^{-4}	2.193×10^{-6}

For family income, restricted activity days, and hospital days, the bias of VAR($R1S$) also decreases with increasing sample size. For the first two variables mentioned, the bias is positive in design I, but in design III, it is negative. For physician visits, the bias in design II is slightly greater than the bias in design I, while for design III, it is at its smallest value. VAR($R1S$)'s bias shows a different trend for proportion seeing a physician. The bias is -7.095×10^{-6} in design I, which increases to 1.581×10^{-7} in design II and then drops to 4.069×10^{-6} in design III. This is the only example for VAR($R1S$) and VAR ($R1H$) for which the bias first is negative and then becomes positive.

VAR($L1$) exhibits a different pattern. The biases for physician visits and proportion seeing a physician are negative for design I. These biases decrease across the designs. However, the method goes from underestimating to overestimating the variance. The bias is the smallest in design II for family income and hospital days.

Comparing the biases of the three methods does not yield a consistent pattern of one of the methods having less bias than the other. The important fact is that for all three estimators, the bias is small and tolerable. However, turning to the variances of the methods in table 6, more of a trend is found. The variances of the variance estimates are:

$$\text{VAR}[\text{VAR}(R1H)] = \frac{1}{900} \times \sum_{j=1}^{900} \left\{[\text{VAR}(R1H)]_j - \overline{\text{VAR}(R1H)}\right\}^2$$

$$\text{VAR}[\text{VAR}(R1S)] = \frac{1}{900} \times \sum_{j=1}^{900} \left\{[\text{VAR}(R1S)]_j - \overline{\text{VAR}(R1S)}\right\}^2$$

$$\text{VAR}[\text{VAR}(L1)] = \frac{1}{900} \times \sum_{j=1}^{900} \left\{[\text{VAR}(L1)]_j - \overline{\text{VAR}(L1)}\right\}^2$$

The first point to consider when examining the variance of estimators is to check to make sure the variance does decrease with increasing sample size. This is true for all three methods.

Table 6. The variance of the $R1H$, $R1S$, and $L1$ variance estimators of the sample estimator $\hat{R}_1$ of the population parameter R

Variance estimator	Variable				
	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R1H$)	1.144×10^{9}	2.052	2.859×10^{-3}	2.428×10^{-3}	1.733×10^{-8}
VAR($R1S$)	1.111×10^{9}	1.976	2.763×10^{-3}	1.902×10^{-3}	1.579×10^{-8}
VAR($L1$)	1.062×10^{9}	1.487	2.380×10^{-3}	2.833×10^{-4}	1.265×10^{-8}
	Design II				
VAR($R1H$)	3.217×10^{8}	0.521	8.306×10^{-4}	1.199×10^{-4}	3.471×10^{-9}
VAR($R1S$)	3.201×10^{8}	0.505	8.274×10^{-4}	1.042×10^{-4}	3.307×10^{-9}
VAR($L1$)	3.258×10^{8}	0.460	8.250×10^{-4}	5.651×10^{-5}	3.338×10^{-9}
	Design III				
VAR($R1H$)	1.047×10^{8}	0.140	8.271×10^{-5}	1.319×10^{-5}	1.104×10^{-9}
VAR($R1S$)	1.036×10^{8}	0.137	8.132×10^{-5}	1.288×10^{-5}	1.089×10^{-9}
VAR($L1$)	1.048×10^{8}	0.131	7.959×10^{-5}	1.135×10^{-5}	1.086×10^{-9}

The variance of the VAR($R1H$) is always larger than the variance of VAR($R1S$). This provides empirical evidence that $R1S$ is more precise than the other forms of replication variance estimators. With three exceptions—family income in designs II and III and proportion seeing a physician in design II—the variance of the VAR($L1$) method is less than the variance of VAR($R1S$).

It is preferable to judge the merits of the three methods on the basis of the mean square error, which takes into consideration both properties of bias and variance. The ideal estimator would be one that produces minimum mean square error for all values of the parameter $\sigma^2_{\hat{R}_1}$. Perhaps such an estimator does not exist, but the empirical results here provide evidence on the behavior of the mean square error for the $R1H$, $R1S$, and $L1$ estimator methods. The formulas for the mean square errors (MSE) are

$$\begin{aligned} \text{MSE}[\text{VAR}(R1H)] &= \text{VAR}[\text{VAR}(R1H)] + \text{bias}^2[\text{VAR}(R1H)] \\ &= \text{VAR}[\text{VAR}(R1H)] + [\overline{\text{VAR}(R1H)} - s^2_{\hat{R}_1}]^2 \end{aligned}$$

$$\begin{aligned} \text{MSE}[\text{VAR}(R1S)] &= \text{VAR}[\text{VAR}(R1S)] + \text{bias}^2[\text{VAR}(R1S)] \\ &= \text{VAR}[\text{VAR}(R1S)] + [\overline{\text{VAR}(R1S)} - s^2_{\hat{R}_1}]^2 \end{aligned}$$

and

$$\begin{aligned} \text{MSE}[\text{VAR}(L1)] &= \text{VAR}[\text{VAR}(L1)] + \text{bias}^2[\text{VAR}(L1)] \\ &= \text{VAR}[\text{VAR}(L1)] + [\text{VAR}(L1) - s^2_{\hat{R}_1}]^2 \end{aligned}$$

As can be observed in table 7, the mean square errors for the methods for each variable decrease with increasing sample size. When looking at the values for the procedures pairwise, the results indicate that VAR($R1S$) always has a smaller mean square error than VAR ($R1H$), which is easily explained by the less

Table 7. The mean square error of the $R1H$, $R1S$, and $L1$ variance estimators of the sample estimator $\hat{R}_1$ of the population parameter R

Variance estimator	Variable				
	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R1H$)	1.159×10^9	2.127	2.866×10^{-3}	2.442×10^{-3}	1.736×10^{-8}
VAR($R1S$)	1.128×10^9	2.053	2.770×10^{-3}	1.916×10^{-3}	1.584×10^{-8}
VAR($L1$)	1.072×10^9	1.499	2.396×10^{-3}	2.883×10^{-4}	1.510×10^{-8}
	Design II				
VAR($R1H$)	3.239×10^8	0.522	8.379×10^{-4}	1.236×10^{-4}	3.731×10^{-9}
VAR($R1S$)	3.228×10^8	0.507	8.323×10^{-4}	1.074×10^{-4}	3.557×10^{-9}
VAR($L1$)	3.259×10^8	0.485	7.653×10^{-4}	5.651×10^{-5}	3.386×10^{-9}
	Design III				
VAR($R1H$)	1.055×10^8	0.140	8.387×10^{-5}	1.320×10^{-5}	1.122×10^{-9}
VAR($R1S$)	1.044×10^8	0.138	8.240×10^{-5}	1.288×10^{-5}	1.106×10^{-9}
VAR($L1$)	1.064×10^8	0.135	8.010×10^{-5}	1.145×10^{-5}	1.091×10^{-9}

variability of VAR($R1S$). The other finding is that VAR($L1$) has a smaller mean square error than VAR($R1S$) aside from the variable family income in designs II and III. The mean square error is less even though VAR($L1$) did not always have the smaller bias. For the only two cases where the mean square error of VAR($L1$) is not lower, the variance of VAR($L1$) was greater. In the other situation, proportion seeing a physician in design II, of greater variance for the $L1$ method, the difference in the bias was large enough to cause VAR($L1$) to have the smaller mean square error.

The behavior of the replication and linearization methods for the properties of bias, variance, and mean square error is very good, which implies the methods are yielding acceptable variance estimates. However, there are other properties to consider when deciding on a variance method; one of these is examined next.

Normality of Standardized Estimators

As mentioned previously, scientists are increasingly concerned with making inferences from data collected in scientific sample surveys rather than just calculating descriptive statistics. Ideally, an investigator would like to construct confidence intervals for population parameters. The statement that the average annual income per person lies between \$7,500 and \$9,400 with 95 percent confidence is more useful than a point estimate of the average income. Such a statement using the normal distribution could be made if one is willing to assume that the variable is sufficiently close to being normally distributed and that the departure from random sampling is not a severe limitation.

The objective here is to study empirically the ratio estimate minus its expected value divided by its estimated standard error. The analysis consists of computing the proportion of times the statistic falls into certain regions. By this procedure, the applicability of confidence statements based on the normal distribution can be approached.

In this section, only the statistics for which VAR($R1H$), VAR($R1S$), and VAR($L1$) were used as the variance estimates will be discussed. These ratios are

$$\frac{\hat{R}_1 - E(\hat{R}_1)}{\sqrt{\text{VAR}(R1H)}},$$

$$\frac{\hat{R}_1 - E(\hat{R}_1)}{\sqrt{\text{VAR}(R1S)}},$$

and

$$\frac{\hat{R}_1 - E(\hat{R}_1)}{\sqrt{\text{VAR}(L1)}}$$

where

$$E(\hat{R}_1) = \frac{1}{900} \sum_{j=1}^{900} \hat{R}_{1j}$$

Such ratios for each estimate in every sample and the proportion of times the ratio fell within the regions—(-1.000, 1.000), (-1.000, 0), (0, 1.000), (-1.645, 1.645), (-1.645, 0), (0, 1.645), (-1.960, 1.960), (-1.960, 0), (1.960, 0), and (-2.576, 2.576)—were computed. These regions were chosen for purposes of comparison with the normal distribution. The proportion of area under the normal curve for the four symmetric intervals are 0.6827, 0.90, 0.95, and 0.99, respectively.

The proportion of times the ratio fell within the limits is given in tables 8-10. Considering just the symmetric intervals, one immediately notices that intervals based on this type of statistic usually have a lower confidence than the normal level, but the differences are generally very small. There are instances in which the proportion of time the statistic based on these different variance estimates fell within the limits exceeded the corresponding normal values. These results are good news, indeed, because coupled with the findings of Frankel[28] they justify the type of inferences consumers of data collected in multistage complex sample surveys often wish to make. For example, a health planner concerned with the use or expansion of health care facilities would most likely need to draw an inference from the estimated proportion of the population who saw a physician at least once in the past 12 months. The data indicate that such an inference based on the normal distribution is really not a bad approximation and could be used with reasonable confidence.

A word of caution about these findings is appropriate. Inspection of the one-sided inter-

Table 8. Proportion of times the sample estimate minus its expected value divided by *R1H* estimate of standard error is within the stated limits

Variable	Limits									
	±1.000	−1.000, 0	0, 1.000	±1.645	−1.645, 0	0, 1.645	±1.960	−1.960, 0	0, 1.960	±2.576
	Design I									
Family income	0.6789	0.3411	0.3378	0.8833	0.4500	0.4333	0.9356	0.4744	0.4611	0.9800
Restricted activity days	0.6778	0.3322	0.3456	0.8811	0.4411	0.4400	0.9367	0.4711	0.4656	0.9789
Physician visits	0.6444	0.2989	0.3456	0.8744	0.4300	0.4444	0.9233	0.4633	0.4600	0.9822
Hospital days	0.6556	0.2933	0.3622	0.8622	0.4156	0.4467	0.9122	0.4489	0.4633	0.9578
Proportion seeing a physician	0.6567	0.3433	0.3133	0.8689	0.4400	0.4289	0.9300	0.4633	0.4667	0.9811
	Design II									
Family income	0.6944	0.3567	0.3378	0.8833	0.4467	0.4367	0.9389	0.4733	0.4656	0.9744
Restricted activity days	0.6356	0.2956	0.3400	0.8844	0.4356	0.4489	0.9300	0.4622	0.4678	0.9778
Physician visits	0.6733	0.3244	0.3489	0.8967	0.4478	0.4489	0.9367	0.4744	0.4622	0.9811
Hospital days	0.6722	0.3233	0.3489	0.8733	0.4367	0.4367	0.9289	0.4722	0.4567	0.9767
Proportion seeing a physician	0.6833	0.3489	0.3344	0.9089	0.4589	0.4500	0.9522	0.4789	0.4733	0.9822
	Design III									
Family income	0.6656	0.3189	0.3467	0.8656	0.4167	0.4489	0.9278	0.4522	0.4756	0.9767
Restricted activity days	0.6511	0.3322	0.3189	0.8700	0.4456	0.4244	0.9244	0.4733	0.4511	0.9756
Physician visits	0.6844	0.3333	0.3511	0.8989	0.4456	0.4533	0.9456	0.4678	0.4778	0.9822
Hospital days	0.6789	0.3422	0.3367	0.8789	0.4467	0.4322	0.9244	0.4722	0.4522	0.9656
Proportion seeing a physician	0.6900	0.3511	0.3389	0.8756	0.4378	0.4378	0.9222	0.4556	0.4667	0.9833

Table 9. Proportion of times the sample estimate minus its expected value divided by $R1S$ estimate of standard error is within the stated limits

Variable	Limits									
	±1.000	−1.000, 0	0, 1.000	±1.645	−1.645, 0	0, 1.645	±1.960	−1.960, 0	0, 1.960	±2.576
	Design I									
Family income	0.6800	0.3456	0.3344	0.8822	0.4478	0.4344	0.9344	0.4733	0.4611	0.9789
Restricted activity days	0.6722	0.3278	0.3444	0.8811	0.4356	0.4456	0.9378	0.4700	0.4678	0.9789
Physician visits	0.6556	0.3067	0.3489	0.8733	0.4300	0.4433	0.9244	0.4622	0.4622	0.9778
Hospital days	0.6611	0.2956	0.3656	0.8644	0.4178	0.4467	0.9156	0.4533	0.4622	0.9578
Proportion seeing a physician	0.6578	0.3433	0.3144	0.8656	0.4356	0.4300	0.9322	0.4667	0.4656	0.9811
	Design II									
Family income	0.6911	0.3567	0.3344	0.8822	0.4456	0.4367	0.9378	0.4733	0.4644	0.9756
Restricted activity days	0.6344	0.2956	0.3389	0.8833	0.4333	0.4500	0.9311	0.4644	0.4667	0.9756
Physician visits	0.6722	0.3233	0.3489	0.8978	0.4500	0.4478	0.9367	0.4733	0.4633	0.9778
Hospital days	0.6800	0.3244	0.3556	0.8744	0.4389	0.4356	0.9344	0.4744	0.4600	0.9767
Proportion seeing a physician	0.6767	0.3456	0.3311	0.9133	0.4611	0.4522	0.9522	0.4800	0.4722	0.9822
	Design III									
Family income	0.6678	0.3200	0.3478	0.8667	0.4189	0.4478	0.9256	0.4500	0.4756	0.9778
Restricted activity days	0.6511	0.3311	0.3200	0.8689	0.4456	0.4233	0.9222	0.4722	0.4500	0.9778
Physician visits	0.6878	0.3356	0.3522	0.8978	0.4422	0.4556	0.9422	0.4656	0.4767	0.9822
Hospital days	0.6789	0.3422	0.3367	0.8789	0.4478	0.4311	0.9267	0.4722	0.4544	0.9644
Proportion seeing a physician	0.6900	0.3511	0.3389	0.8778	0.4389	0.4389	0.9256	0.4589	0.4667	0.9833

vals indicates that they are not symmetric. Thus, any statements about one-sided intervals cannot be made, and confidence intervals computed are not of minimum length. An area of further study could be the direct corroboration of the skewness and leptokurtosis noted in the tables by computing and examining the third and fourth sample moments.

Influence of Poststratification

Another purpose of the study is to determine the effect the estimation component poststratification has on variances. The reason for adjusting is to lower the variance. Normally, to reduce variance a larger sample is drawn, but if the same thing can be accomplished by the estimation procedure, there can be a savings in cost by using a smaller sample size, unless, of course, the savings in field costs are offset by costs of implementing the estimation procedure. Variance reduction has been demonstrated primarily using simplified methods and theoretical analysis, while the studies of Simmons and Bean[20] and Banks and Shapiro[8] gave empirical evidence. This investigation examines the effect for the particular design used here.

The VAR($R1H$) and VAR($R1C$) schemes call for the poststratification factors to be computed within each half-sample and each complement half-sample, while for the methods VAR($R2H$) and VAR($R2C$), the parent sample adjustment factors are used. (See appendix II for formulas for VAR($R2H$) and VAR($R2C$).) The orthog-

Table 10. Proportion of times the sample estimate minus its expected value divided by $L1$ estimate of standard error is within the stated limits

Variable	±1.000	−1.000, 0	0, 1.000	±1.645	−1.645, 0	0, 1.645	±1.960	−1.960, 0	0, 1.960	±2.576
	Limits									
	Design I									
Family income . .	0.6733	0.3400	0.3333	0.8622	0.4378	0.4244	0.9211	0.4667	0.4544	0.9722
Restricted activity days . . .	0.6433	0.3189	0.3244	0.8611	0.4244	0.4367	0.9244	0.4667	0.4578	0.9711
Physician visits .	0.6211	0.2867	0.3344	0.8489	0.4222	0.4267	0.9044	0.4522	0.4522	0.9711
Hospital days . .	0.6200	0.2844	0.3356	0.8344	0.4011	0.4333	0.8933	0.4411	0.4522	0.9600
Proportion seeing a physician . .	0.6189	0.3211	0.2978	0.8378	0.4244	0.4133	0.9022	0.4511	0.4511	0.9711
	Design II									
Family income .	0.6811	0.3522	0.3289	0.8833	0.4444	0.4389	0.9289	0.4667	0.4622	0.9744
Restricted activity days . . .	0.6178	0.2867	0.3311	0.8644	0.4211	0.4433	0.9211	0.4567	0.4644	0.9711
Physician visits .	0.6533	0.3167	0.3367	0.8900	0.4467	0.4433	0.9356	0.4744	0.4611	0.9767
Hospital days . .	0.6511	0.3156	0.3356	0.8533	0.4256	0.4278	0.9200	0.4678	0.4522	0.9700
Proportion seeing a physician . .	0.6544	0.3378	0.3167	0.9022	0.4522	0.4500	0.9467	0.4744	0.4722	0.9800
	Design III									
Family income .	0.6644	0.3167	0.3478	0.8600	0.4133	0.4467	0.9256	0.4489	0.4767	0.9767
Restricted activity days . . .	0.6400	0.3256	0.3144	0.8611	0.4411	0.4200	0.9156	0.4700	0.4456	0.9756
Physician visits .	0.6811	0.3311	0.3500	0.8956	0.4444	0.4511	0.9378	0.4622	0.4756	0.9822
Hospital days . .	0.6689	0.3322	0.3367	0.8733	0.4456	0.4278	0.9233	0.4744	0.4489	0.9644
Proportion seeing a physician . .	0.6867	0.3489	0.3378	0.8722	0.4367	0.4356	0.9244	0.4567	0.4689	0.9844

onal pattern used for the methods was the same, producing identical composition. The ratio of the variance estimator secured from VAR($R1H$) to the estimator produced by VAR($R2H$) does measure the relative impact of poststratification. The average of these ratios was calculated. The ratios are defined to be

$$\text{average ratio of variances for method } i = \frac{1}{900}\sum_{j=1}^{900}\frac{[\text{VAR}(R1i)]_j}{[\text{VAR}(R2i)]_j}$$

where

$i = H, C, S.$

Only the third replication scheme is discussed. Computing the poststratification weights in each half-sample reduces variance estimates for family income and physician visits (see the ratio of VAR($R1S$) to VAR($R2S$) in table 11. This is true for all three designs. A reduction is achieved for all the remaining variables in design III. These data provide evidence that poststratification does improve precision.

SUMMARY

Introduction

Increasing use of scientific survey sampling has led to the development of a wide variety of techniques. These have been produced in re-

Table 11. The relative impact of poststratification on the variance estimates

Variable	Average value of the ratio of		
	VAR(*R1H*) / VAR(*R2H*)	VAR(*R1C*) / VAR(*R2C*)	VAR(*R1S*) / VAR(*R2S*)
	Design I		
Family income	0.833	0.833	0.833
Restricted activity days	1.030	1.030	1.031
Physician visits	1.001	1.000	1.000
Hospital days	1.035	1.036	1.035
Proportion seeing a physician	1.034	1.026	1.030
	Design II		
Family income	0.840	0.844	0.842
Restricted activity days	1.002	1.005	1.003
Physician visits	0.973	0.973	0.973
Hospital days	1.015	1.004	1.009
Proportion seeing a physician	0.985	0.984	0.984
	Design III		
Family income	0.851	0.850	0.850
Restricted activity days	0.980	0.982	0.981
Physician visits	0.961	0.961	0.961
Hospital days	0.967	0.966	0.967
Proportion seeing a physician	0.975	0.973	0.974

sponse to the need to sample populations that are geographically scattered, requiring large PSU's. The use of poststratification to adjust the distribution of the sample for certain demographic characteristics has produced further methodologic refinements. Such features justify the generic term "complex multistage probability samples" to describe the methods. Because of the complexity of these samples, variance formulas for estimators are themselves complicated, and often only approximate expressions can be obtained. Also, since the criteria of simple random sampling, independence, and normality are not met, classical statistical procedures must be examined to determine their applicability in such an environment. This investigation has employed an empirical approach to consider various problems of estimating variances and making inferences for complex samples.

By the procedure of Monte Carlo sampling from a completely specified universe, the following points were investigated:

1. Behavior of the two general variance estimator methods: linearization and replication
2. Distribution of the ratio of an estimated mean minus its expected value divided by its standard error with the normal distribution
3. Study of the bias of two estimators
4. Investigation of a simpler variance estimator as an approximation to the correct replication and linearization procedures
5. Measurement of the impact of poststratification on variance estimates produced by the two general methods

The study considered a particular sample design and estimation process. The design was complex, involving stratification and two stages of clustering with poststratification. The results are for 900 samples drawn for three different sample sizes and, thus, are based on a considerable body of evidence.

Behavior of Linearization and Replication

The variance of a poststratified ratio estimator was estimated by the replication and linearization methods. Three different variance estimators were computed from the replication method. For each of 900 samples drawn for three sample sizes, two ratio estimates and their variances were computed. The variance estimates produced by these techniques VAR($R1H$), VAR($R1S$), and VAR($L1$) were compared on the bases of bias, variance, and mean square error.

For the sample design and estimation procedure employed here, both the replication and linearization methods provide generally adequate variance estimates. Almost no severe abnormalities were observed. The biases of the methods are satisfactory, and the measures of total mean square error for the methods are adequately small with one exception. The variance of the VAR($L1$) method is generally less than the variance of VAR($R1S$), especially for the variable hospital days.

Normality of Standardized Estimators

The distribution of the ratio

$$\frac{\text{sample estimate} - E(\text{sample estimate})}{\sqrt{\text{VAR(sample estimate)}}}$$

was studied. When the sample design and estimation procedure result in a complex survey, the exact distribution of this ratio is unknown. For this study, such ratios were computed for each estimator, VAR($R1H$), VAR($R1S$), and VAR($L1$). To examine confidence interval statements, the proportion of times the ratio fell within stated limits was computed. The symmetric limits chosen were the normal deviates for the 99th, 95th, 90th, and 68.27th quantiles of a normal distribution. Also calculated was the proportion of times the ratio fell in the one-sided intervals (-1.000, 0) (0, +1.000), (-1.645, 0) (0, 1.645), (-1.960, 0), and (0, 1.960).

Proportions for the symmetric intervals with VAR($R1S$), VAR($R1H$), and VAR($L1$) as the variance estimator were exceedingly close to corresponding normal values. For the most part, the proportions for the three methods were smaller than the proportions for the normal distribution. Examination of the one-sided intervals showed that the proportions on either side of zero are not the same. This fact and the indication of leptokurtosis provided evidence that the distributions of the ratios are not symmetric.

This body of empirical data indicates that approximate interval estimates based on normal distribution theory can be made. Other important points are that (1) the true confidence level is somewhat lower than the nominal level of the normal, (2) one-sided intervals cannot be constructed, and (3) because of asymmetry, these intervals are not minimum length.

Comparison of Two Ratio Estimators

Sampling theory states that ratio estimators are biased.[2,4] In the present study, two estimators, distinguished by the level of poststratification, were investigated: $\hat{R}_1$ was adjusted to the universe totals of a class, and $\hat{R}_2$ was a combined ratio estimator with the poststratification adjustment at the region level. These estimators were calculated for each of the five variables for every selected sample.

The bias and relative bias were reported. The differential bias between $\bar{r}_1$ (average of replicate and complement estimators of R) and $\hat{R}_1$ was also determined using the first replication equation. None of the relative biases was greater than 1 percent. Thus, these ratio estimators are essentially unbiased for the sample sizes studied.

Approximate Variance Estimators

Two computationally simpler methods were included in the study to determine whether one of them could be used to produce approximate variance estimates of $\hat{R}_1$. The first method was a replication method in which the parent sample adjustment factors were used in place of separate factors for each half-sample. The second method produced an overall variance estimate for $\hat{R}_1$ by first estimating the variances of $\hat{R}_{ENC,1}$ and $\hat{R}_{SW,1}$ and then combining the two correctly. Included in this discussion were the results of estimating variances correctly for $\hat{R}_2$.

Only the first approximation VAR($R2S$) can be seriously considered for possible use. This method was more biased than the correct procedure, but the mean square errors were within tolerable bounds. The proportions produced when this variance estimator was used do not have large deviations from the normal values. However, the direction of the deviations varied more than for the VAR($R1S$) method.

The behavior of VAR($R4S$), which is a linear combination of subuniverse estimators, was adequate. The difference between this method and the approximate one, VAR($R3S$), is that VAR ($R4S$) is the appropriate variance estimator for the ratio estimator $\hat{R}_2$. The ratio estimator $\hat{R}_2$ is poststratified to the regional distribution. Therefore, the variances are weighted averages of the variance estimators of the region estimators. The bias, variance, and mean square error were found to be acceptable and not to be extremely large. The findings on the ratio of the sample estimator minus its expected value divided by its standard error were that for VAR($R4S$) the empirical proportions for the symmetric intervals compared favorably to the proportions for the normal distributions.

Influence of Poststratification

Theory states that the benefits derived from adjusting sample results to population totals for selected characteristics lie in improvement of the precision of the ratio estimator by insuring that domains of study have the same distribution in the sample as in the population. Assessment of the effect of poststratification on variances was a natural byproduct of the calculations done for the other methods. The estimates of variances produced by the first replication method were compared with those secured from the second equation. In the first scheme, poststratification factors were computed within each half-sample and each complement using half the sample data, while the parent sample factors were applied when making these estimates in the second scheme.

VAR($R1S$) showed a reduction in variance for family income in design I; for family income, physician visits, and proportion seeing a physician in design II; and for all the variables in design III.

Conclusions

The outstanding finding was that both the replication method and the linearization method are highly satisfactory methods for estimating variances of poststratified ratio estimators obtained from complex probability surveys of the type studied here. Features of the sample design employed were unequal sampling fractions, stratification, and two stages of clustering. Since the methods yield adequate variance estimates, they provide the means for drawing valid inferences with known precision from such surveys. For these data, the linearization variance estimation method had slightly lower variance and mean square error than the replication method. Neither method showed substantial bias.

The proportions based on the ratios of the sample estimate minus its expected value divided by its standard error as estimated by the replication method were closer to the corresponding normal values than the proportions with the linearization method as the variance estimator. However, the ratios with the $L1$ variance estimator had a more consistent pattern than the ratios with the standard error estimated by replication method. The consistency was evidenced by the difference between empirical proportions and the corresponding ones for the normal distribution getting smaller as sample size increased.

The applicability of confidence intervals by using normal distribution theory has been shown to be adequate. The true level of confidence is usually lower than the confidence stated, which should be brought out when making inferences. Another item that should be mentioned is that the distributions of the ratios are not symmetric. Thus, one-sided intervals are not valid and the intervals constructed are not of minimum length.

The biases of the ratio poststratified estimators were negligible for the sample sizes studied.

The VAR($R2S$) estimator proved to be an excellent approximation, and its use is recommended. However, there is a tradeoff involved when using an approximation. The gain in simpler formulas may be offset by larger biases. Also, the proportions produced when the VAR($R2S$) was used to estimate standard errors varied widely in the direction of its deviation from the proportions of the normal distribution. The investigator should consider that consumers of the statistics may make such inferences whether or not he does. Then each must weigh the losses and the consequences when deciding about a variance estimation method.

The effect of the estimation component poststratification was to lower variance and mean square error. This reduction is important in making inferences from complex sample surveys and can be converted into cost savings.

Relevancy to the HIS

The striking feature of this investigation has been to provide empirical evidence that data collected in complex multistage probability surveys such as the HIS can be used for analytical purposes. Variance estimates, which are essential in analyses, can be calculated by using either the replication or the linearization variance estimator method. Having adequate variances opens avenues of further research in developing testing techniques and other refined analyses, perhaps along the lines of multivariate procedures, comparable to classical ones.

Equally rewarding as finding that suitable variance estimates can be computed was the result that tests and confidence intervals can be approximated by normal distribution theory. This allows one to draw inferences about the morbidity estimates published by the HIS.

The HIS has now a choice of either of the two methods to use in computations. The variability of the linearization method was slightly lower than that of the replication method, but neither is very biased. The linearization equations can easily be interpreted to obtain estimates of components of variance, which are crucial in planning such surveys. Another factor to consider when choosing a variance estimator method is that the replication technique is easily applicable to the estimation of variances for many types of estimators without deriving new theory or new computer programs for each estimator. This is not true for the linearization method.

———— ○○○ ————

REFERENCES

[1]Kish, Leslie: Confidence intervals for clustered samples. *Am Soc. Rev.* 22(2):154-165, Apr. 1957.

[2]Kish, Leslie: *Survey Sampling.* New York. John Wiley & Sons, Inc., 1965.

[3]Deming, William E.: *Some Theory of Sampling.* New York. John Wiley & Sons, Inc., 1950.

[4]Hansen, M. H., Hurwitz, W. N., and Madow, W. G.: *Sample Survey Methods and Theory.* Vol. 1. New York. John Wiley & Sons, Inc., 1953.

[5]"Student." (W. S. Gossett): The probable error of a mean. *Biometrika* 6(Pt. I):1-25, Mar. 1908.

[6]National Center for Health Statistics: Replication: An approach to the analysis of data from complex surveys. *Vital and Health Statistics.* Series 2, No. 14. DHEW Pub. No. (HSM) 73-1269. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Apr. 1966.

[7]U.S. Bureau of the Census: The Current Population Survey: A report on methodology. Tech. Paper No. 7. Washington. U.S. Government Printing Office, 1963.

[8]Banks, Martha J., and Shapiro, Gary M.: Variances of the Current Population Survey, including within- and between-PSU components and the effect of the different stages of estimation. *Proc. Soc. Statist. Sec. Am. Statist. A.*, 1971, 40-49.

[9]Gurney, Margaret: The variance of the replication method for estimating variances for the CPS sample design. Washington. U.S. Bureau of the Census, 1962. (Dittoed memorandum.)

[10]National Center for Health Statistics: Cycle I of the Health Examination Survey: Sample and Response, United States, 1960-1962. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 11-No. 1. Public Health Service. Washington. U.S. Government Printing Office, 1964.

[11]McCarthy, Philip J., Simmons, Walt R., and Losee, Garrie J.: Replication techniques for estimating the variances from complex surveys. Paper presented at the Joint Session of the Joint Epidemiology and Statistics Section of the American Public Health Association, Chicago, Ill., Oct. 18, 1965.

[12]Simmons, Walt R., and Baird, James T., Jr.: Pseudo-replication in the NCHS Health Examination Survey. *Proc. Soc. Statist. Sec. Am. Statist. A.*, 1968, 19-30.

[13]Plackett, R. L., and Burman, J. P.: The design of optimum multifactorial experiments. *Biometrika* 33(Pt. IV):305-325, June 1946.

[14]Lee, Kok-huat: Partially balanced designs for half-sample replication method of variance estimation. *J. Am. Statist. A.* 67(338):324-334, June 1972.

[15]National Center for Health Statistics: Pseudo-replication: Further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics.* Series 2, No. 31. DHEW Pub. No. (HSM) 73-1270. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Jan. 1969.

[16]McCarthy, Philip J.: Pseudo-replication: Half samples. *Rev. International Statist. Inst.*, 37:239-264, 1969.

[17]Kish, Leslie, and Frankel, Martin: Balanced repeated replications for analytical statistics. *Proc. Soc. Statist. Sec. Am. Statist. A.*, 1968, 2-10.

[18]Kish, Leslie, and Frankel, Martin R.: Balanced repeated replication for standard errors. *J. Am. Statist. A.*, 65:1071-1094, 1970.

[19]Levy, Paul S.: A comparison of balanced half-sample replication and jackknife estimators of the variances of ratio estimates in complex sampling with two primary sampling units per stratum. Rockville, Md. National Center for Health Statistics, 1971. (Dittoed report.)

[20]Simmons, Walt R., and Bean, Judy A.: Impact of design and estimation components on inference. *New Developments in Survey Sampling.* (N. L. Johnson and H. Smith, Jr., eds.) New York. Wiley-Interscience, 1969, pp. 601-628.

[21]Keyfitz, Nathan: Estimates of sampling variance where two units are selected from each stratum. *J. Am. Statist. A.*, 52(280):503-510, Dec. 1957.

[22]Hansen, M. H., Hurwitz, W. N., and Madow, W. G.: *Sample Survey Methods and Theory.* Vol. 2. New York. John Wiley & Sons, Inc., 1953.

[23]Tepping, Benjamin J.: Variance estimation in complex surveys. *Proc. Soc. Statist. Sec. Am. Statist. A.*, 1968, 11-18.

[24]National Center for Health Statistics: Estimation and sampling variance in the Health Interview Survey. *Vital and Health Statistics.* Series 2, No. 38. DHEW Pub. No. (HRA) 74-1288. Health Resources Administration. Washington. U.S. Government Printing Office, June 1970.

[25]Shapiro, Gary M.: Keyfitz method of estimating variance and its application to the Current Population Survey. Washington. U.S. Bureau of the Census, 1966. (Dittoed report.)

[26]Waltman, Henry: Variance estimation formulas to be used in computing total NSR PSU variances for the 449-area design. Washington. U.S. Bureau of the Census, 1968. (Dittoed report.)

[27]Fellegi, I. P., and Gray, G. B.: Sampling errors in periodic surveys. *Proc. Soc. Statist. Sec. Am. Statist. A.*, 1971, 28-35.

[28]Frankel, Martin R.: *Inference from Survey Samples: An Empirical Investigation.* Ann Arbor. Institute for Social Research, University of Michigan, 1971.

[29]U.S. National Health Survey: The statistical design of the Health Household-Interview Survey. *Health Statistics.* PHS Pub. No. 584-A2. Public Health Service. Washington. U.S. Government Printing Office, July 1958.

[30]Bean, Judy A.: Behavior of replication and linearization variance estimators for complex multistage probability samples. Unpublished doctoral dissertation. University of Texas, 1973.

APPENDIX I

EXTENT OF BIASNESS OF TWO RATIO ESTIMATORS

Introduction

Another objective of the study was to determine the extent of biasness of two ratio estimators. The formulas used and the results follow.

Estimators

The two estimators of a statistic follow:

1. $\hat{R}_1$—poststratification at universe level. The formula is given in the section Estimators and Variance Estimators.
2. $\hat{R}_2$—poststratification at region level

$$\hat{R}_2 = \frac{x''_{\text{ENC},2} + x''_{\text{SW},2}}{y}$$

$$= \frac{\sum_{a=1}^{24} x'_{\text{ENC},a} \frac{y_{\text{ENC},a}}{y'_{\text{ENC},a}} + \sum_{a=1}^{24} x'_{\text{SW},a} \frac{y_{\text{SW},a}}{y'_{\text{SW},a}}}{y}$$

where

ENC = East and North Central regions,

SW = South and West regions,

$\hat{R}_2$ = poststratified ratio estimate of the population parameter R,

$$x''_{\text{ENC},2} = \sum_{a=1}^{24} x'_{\text{ENC},a} \frac{y_{\text{ENC},a}}{y'_{\text{ENC},a}},$$

$x'_{\text{ENC},a}$ = the sample estimate for class a in the ENC Region,

$y_{\text{ENC},a}$ = the population size in class a in the ENC Region,

$$x''_{\text{SW},2} = \sum_{a=1}^{24} x'_{\text{SW},a} \frac{y_{\text{SW},a}}{y'_{\text{SW},a}}$$

$x'_{\text{SW},a}$ = the sample estimate for class a in the SW Region,

$y_{\text{SW},a}$ = the population size in class a in the SW Region,

$y'_{\text{SW},a}$ = the sample estimate of population size for class a in the SW Region.

Comparison of Two Ratio Estimators

Ratio estimators are commonly employed in complex multistage sample surveys and usually contain the features of unequal sampling fractions and poststratification. The distributions of these ratio estimators are unknown. In this investigation, two such ratio estimators were employed in order to study the extent of their biasness.

Each estimation procedure included inflation by the reciprocal of the selection probability and poststratification by 24 ethnic-sex-age classes. This probability of selection is the product of the probabilities of selection from each step of selection: PSU and segment. The level at which poststratification was performed differentiates the estimators. The $\hat{R}_1$ estimator was adjusted to a universe level; $\hat{R}_2$ was adjusted to the distribution of subgroups within the ENC Region and within the SW Region.

Ratio estimators are biased estimators. The extent of this bias has been studied both theoretically and empirically. Since these ratio estimates were calculated for each of five variables for 900 samples for three sample sizes, sample empirical evidence was available to investigate the magnitude of the bias of the poststratified ratio estimators employed here. The universe was completely specified, so the parameter R was known. A measure of the bias is

$$\text{bias }(\hat{R}_i) = \bar{\hat{R}}_i - R$$

where

R = parameter value,

$$\bar{\hat{R}}_i = \frac{1}{900}\sum_{j=1}^{900} \hat{R}_{ij}, \quad i=1,2$$

Table I shows that neither of the two estimators consistently had the smaller bias. In design I, with the exception of the variable proportion seeing a physician, $\hat{R}_1$ has the smaller value. For the other two designs, $\hat{R}_1$ does not compare as well, and in fact, only exhibits the smaller bias for the variable proportion seeing a physician in design II and for hospital days in III.

The relative bias of the estimators used for another evaluation of the estimators is defined to be

$$\text{relative bias }(\hat{R}_i) = \frac{\bar{\hat{R}}_i - R}{R}$$

where

R = the population parameter,

$\bar{\hat{R}}_i$ = the sample mean of the estimates produced by ratio estimator i,

$$= \frac{1}{900}\sum_{j=1}^{900} \hat{R}_{ij}, \quad i=1,2$$

Table II shows the values for the two estimates for each variable and design. The relative bias is always less than 1 percent for any of the

Table I. The bias of two ratio estimators

Variable	Estimator $\hat{R}_1$	Estimator $\hat{R}_2$
	Design I	
Family income	-0.0047	-0.0124
Restricted activity days	0.0049	0.0021
Physician visits	-0.0135	-0.0197
Hospital days	0.0097	0.0082
Proportion seeing a physician	0.0059	-0.0035
	Design II	
Family income	0.0055	0.0047
Restricted activity days	-0.0009	-0.0004
Physician visits	-0.0119	-0.0112
Hospital days	0.0016	0.0009
Proportion seeing a physician	0.0035	0.0041
	Design III	
Family income	-0.0082	-0.0064
Restricted activity days	-0.0012	-0.0010
Physician visits	-0.0210	-0.0198
Hospital days	0.0040	0.0045
Proportion seeing a physician	0.0021	0.0007

Table II. The relative bias of two ratio estimators

Variable	Estimator $\hat{R}_1$	Estimator $\hat{R}_2$
	Design I	
Family income	-0.0006	-0.0015
Restricted activity days	0.0031	0.0014
Physician visits	-0.0029	-0.0040
Hospital days	0.0093	0.0078
Proportion seeing a physician	0.0009	-0.0004
	Design II	
Family income	0.0007	0.0006
Restricted activity days	-0.0006	-0.0003
Physician visits	-0.0025	-0.0024
Hospital days	0.0015	0.0009
Proportion seeing a physician	0.0006	0.0006
	Design III	
Family income	-0.0010	-0.0008
Restricted activity days	-0.0008	-0.0007
Physician visits	-0.0045	-0.0042
Hospital days	0.0039	0.0043
Proportion seeing a physician	-0.0003	0.0001

Table III. Distribution of differential bias of $\bar{r}_1$ and $\hat{R}_1$ as a fraction of the estimated standard error of $\bar{r}_1$ for replication method 1

$\frac{\bar{r}_1 - \hat{R}_1}{\sqrt{\mathrm{VAR}(\bar{r}_1)}}$	Variable				
	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
Less than -.30	4	4	5	7	18
-.30 to -.25	4	6	5	3	12
-.25 to -.20	5	8	2	9	22
-.20 to -.15	18	29	13	10	40
-.15 to -.10	45	49	64	42	97
-.10 to -.05	142	139	129	115	178
-.05 to -.00	345	265	304	291	254
.00 to .05	249	257	267	264	173
.05 to .10	66	98	86	113	74
.10 to .15	17	30	20	27	22
.15 to .20	2	12	3	10	6
.20 to .25	3	2	1	6	1
.25 to .30	0	1	1	3	2
Greater than .30	0	0	0	0	1
	Design II				
Less than -.30	0	0	0	1	0
-.30 to -.25	0	0	0	0	0
-.25 to -.20	0	0	0	1	0
-.20 to -.15	0	2	3	3	2
-.15 to -.10	1	9	7	13	1
-.10 to -.05	24	89	42	78	57
-.05 to .00	327	333	314	318	344
.00 to .05	438	363	423	350	369
.05 to .10	100	93	100	111	115
.10 to .15	7	7	11	17	10
.15 to .20	3	4	0	7	2
.20 to .25	0	0	0	1	0
.25 to .30	0	0	0	0	0
Greater than .30	0	0	0	0	0
	Design III				
Less than -.30	0	0	0	0	0
-.30 to -.25	0	0	0	0	0
-.25 to -.20	0	0	0	0	0
-.20 to -.15	0	0	0	0	0
-.15 to -.10	0	0	0	1	1
-.10 to -.05	4	30	6	25	13
-.05 to .00	231	394	270	328	310
.00 to .05	578	437	577	467	486
.05 to .10	82	39	47	68	82
.10 to .15	5	0	0	9	7
.15 to .20	0	0	0	1	1
.20 to .25	0	0	0	0	0
.25 to .30	0	0	0	0	0
Greater than .30	0	0	0	0	0

estimates and only approaches 1 percent for $\hat{R}_1$ in design I for hospital days. As the sample size increases, the biases should decrease. However, here the relative biases are very small and exhibit no consistent decrease. Also, the relative biases are not always positive or negative, but vary. Neither of the two estimators employed yields the smallest value consistently across the statistics and designs.

McCarthy[15] developed an analytical technique for checking on the bias of the ratio estimate using a byproduct of the replication variance estimation method. This method was employed to provide a further study of the bias of $\hat{R}_1$, since it is the common estimator in complex sample surveys. The replicate estimate $R'_{1,\alpha}$ and the complement estimate $R^*_{1,\alpha}$ and their average $\bar{r}_1$ are each estimates of R secured from half the data rather than the entire sample. Since the bias decreases with increasing sample size, the estimate $\bar{r}_1$ should have greater bias than $\hat{R}_1$. Another measure of the absolute bias of $\hat{R}_1$ is the differential bias between $\bar{r}_1$ and $\hat{R}_1$. If values $\hat{R}_1$ and $\bar{r}_1$ for a set of sample data are nearly the same, it is reasonable to believe the differential bias is small. Thus, the absolute bias of either estimator is also small.

This approach was used to examine the bias of the ratio estimator $\hat{R}_1$. The half-sample estimates R'_1, computed in the VAR($R1H$) scheme, and the estimates R^*_1, calculated in the VAR($R1C$) scheme, and the average of the two estimates across all the half-samples were used. The differential bias of $\bar{r}_1$ and $\hat{R}_1$ was expressed as a fraction of the estimated standard error of $\bar{r}_1$. The formula is

$$\frac{\bar{r}_1 - \hat{R}_1}{\sqrt{\mathrm{VAR}(\bar{r}_1)}}$$

where

$$\bar{r}_1 = \frac{1}{40}\left(\sum_{\alpha=1}^{20} R'_{1,\alpha} + \sum_{\alpha=1}^{20} R^*_{1,\alpha}\right);$$

$R'_{1,\alpha}$ = the αth half-sample estimate of R computed in the $R1H$ variance scheme,

$R^*_{1,\alpha}$ = the αth complement half-sample estimate of R computed in the $R1C$ variance scheme, and

$$\mathrm{VAR}(\bar{r}_1) = \frac{1}{4}\sum_{\alpha=1}^{20}\frac{1}{20}(R'_{1,\alpha} - R^*_{1,\alpha})^2$$

As can be seen in table III, biases appear to be small and to decrease with sample size.

These data indicate that the bias of the two poststratified ratio estimators used in this investigation is small and is not a factor to be concerned about. In the properties studied, none of the two ratio estimators clearly dominated. However, $\hat{R}_1$, with the exception of family income, is perhaps better than $\hat{R}_2$.

——— o o o ———

APPENDIX II

A SIMPLER VARIANCE ESTIMATOR

Introduction

The purpose of this phase of the investigation was to ascertain whether a simpler variance estimator can be used as an approximation to the correct replication and linearization variance estimates of the ratio estimate $\hat{R}_1$. Another aspect explored was to examine an alternative ratio estimator for which correct variance estimates are easier to compute.

Variance Estimators for Estimator 1

The following two replication methods are approximate estimates of the variance of $\hat{R}_1$.

1. VAR($R2H$), VAR($R2C$), and VAR($R2S$)

Here, instead of calculating a new poststratification factor for each half-sample and each complement replicate, the multiplication factor derived in the parent sample was used. Again, the orthogonal pattern in table 4 was used, which means that these replicates were the same as the ones in the variance schemes $R1H$, $R1C$, and $R1S$. The formulas are

$$\text{VAR}(R2H) = \frac{1}{20}\sum_{\alpha=1}^{20}(R'_{2,\alpha} - \hat{R}_1)^2$$

$$\text{VAR}(R2C) = \frac{1}{20}\sum_{\alpha=1}^{20}(R^*_{2,\alpha} - \hat{R}_1)^2$$

and

$$\text{VAR}(R2S) = \frac{1}{2}[\text{VAR}(R2H) + \text{VAR}(R2C)]$$

where

$R'_{2,\alpha}$ = the αth replicate estimate of the parameter with parent poststratification factors,

$R^*_{2,\alpha}$ = the αth complement estimate with parent poststratification factors.

2. VAR($R2H$), VAR($R3C$), and VAR($R3S$)

The variance estimates for the region ratio estimates were calculated and then combined to give another approximation to the variance of $\hat{R}_1$. When the number of strata in a sample design is too large, it may not be feasible to consider employing an orthogonal pattern of size $k \times k$, where k is the necessary number of half-samples required. An illustration of this might be a design which has, say, 400 strata. In this case, an approximation or alternative procedure could be substituted for the correct formula. By computing the estimate in this manner, two orthogonal patterns, one for the ENC Region and another for the SW Region, each requiring less than k replicates, can be employed rather than a single pattern of size $k \times k$.

The ratio estimator R_1 can be expressed as

$$\hat{R}_{\text{ENC},1} = \frac{y_{\text{ENC}}\hat{R}_{\text{ENC},1} + y_{\text{SW}}\hat{R}_{\text{SW},1}}{y}$$

where

$$\hat{R}_{ENC,1} = \frac{x''_{ENC,1}}{y_{ENC}}$$

$$= \frac{\sum_{a=1}^{24} x'_{ENC,a} \frac{y_a}{y'_a}}{y_{ENC}}$$, where the sum is over PSU's in the ENC Region,

$x''_{ENC,1}$ = the poststratified ratio estimate of health characteristic x for the ENC Region,

$$R_{SW,1} = \frac{x''_{SW,1}}{y_{SW}}$$

$$= \frac{\sum_{a=1}^{24} x'_{SW,a} \frac{y_a}{y'_a}}{y_{SW}}$$, where the sum is over PSU's in the SW Region,

$x''_{SW,1}$ = the poststratified ratio estimate of health characteristic x for the SW Region.

Then the variance estimator is

$$\text{VAR}(\hat{R}_1) = \frac{y^2_{ENC}\,\hat{\text{VAR}}(\hat{R}_{ENC,1}) + y^2_{SW}\,\hat{\text{VAR}}(\hat{R}_{SW,1})}{y^2}$$

The two orthogonal patterns used are in tables IV and V. Another feature of this approximation is the level at which poststratification was done for the half-sample estimators. Since the estimator $\hat{R}_1$ is poststratified at a universe level, these factors should be computed within each half-sample at a universe level, in order for the half-sample estimates to be equivalent. However, this cannot be done when two different-sized patterns are used. The solution was to apply the parent factors. Formulas for these variances are

$$\text{VAR}(\hat{R}_{ENC,1} : H) = \frac{1}{12}\sum_{\alpha=1}^{12}(R'_{ENC,\alpha} - \hat{R}_{ENC,1})^2$$

$$\text{VAR}(\hat{R}_{ENC,1} : C) = \frac{1}{12}\sum_{\alpha=1}^{12}(R^*_{ENC,\alpha} - \hat{R}_{ENC,1})^2$$

$$\text{VAR}(\hat{R}_{ENC,1} : S) = \frac{1}{2}[\text{VAR}(\hat{R}_{ENC,1} : H) + \text{VAR}(\hat{R}_{ENC,1} : C)]$$

$$\text{VAR}(\hat{R}_{SW,1} : H) = \frac{1}{12}\sum_{\alpha=1}^{12}(R'_{SW,\alpha} - \hat{R}_{SW,1})^2$$

$$\text{VAR}(\hat{R}_{SW,1} : C) = \frac{1}{12}\sum_{\alpha=1}^{12}(R^*_{SW,\alpha} - \hat{R}_{SW,1})^2$$

$$\text{VAR}(\hat{R}_{SW,1} : S) = \frac{1}{2}[\text{VAR}(\hat{R}_{SW,1} : H) + \text{VAR}(\hat{R}_{SW,1} : C)]$$

$$\text{VAR}(R3H) = y^2_{ENC}[\text{VAR}(\hat{R}_{ENC,1} : H)] + y^2_{SW}[\text{VAR}(\hat{R}_{SW,1} : H]$$

$$\text{VAR}(R3C) = y^2_{ENC}[\text{VAR}(\hat{R}_{ENC,1} : C)] + y^2_{SW}[\text{VAR}(\hat{R}_{SW,1} : C]$$

Table IV. Orthogonal pattern for ENC Region variance estimates[1]

Replicate	Strata in the ENC Region									
	1	2	3	4	5	9	10	11	12	13
1	+	+	–	+	+	+	–	–	–	+
2	+	–	+	+	+	–	–	–	+	–
3	–	+	+	+	–	–	–	+	–	+
4	+	+	+	–	–	–	+	–	+	+
5	+	+	–	–	–	+	–	+	+	–
6	+	–	–	–	+	–	+	+	–	+
7	–	–	–	+	–	+	+	–	+	+
8	–	–	+	–	+	+	–	+	+	+
9	–	+	–	+	+	–	+	+	+	–
10	+	–	+	+	–	+	+	+	–	–
11	–	+	+	–	+	+	+	–	–	–
12	–	–	–	–	–	–	–	–	–	–

[1] "+" denotes +1 and "–" denotes –1.

Table V. Orthogonal pattern for SW Region variance estimates[1]

Replicate	Strata in the SW Region								
	6	7	8	14	15	16	17	18	19
1	–	+	–	–	–	+	+	+	–
2	+	–	+	–	–	–	+	+	+
3	+	+	–	+	–	–	–	+	+
4	–	+	+	–	+	–	–	–	+
5	+	–	+	+	–	+	–	–	–
6	+	+	–	+	+	–	+	–	–
7	+	+	+	–	+	+	–	+	–
8	–	+	+	+	–	+	+	–	+
9	–	–	+	+	+	–	+	+	–
10	–	–	–	+	+	+	–	+	+
11	+	–	–	–	+	+	+	–	+
12	–	–	–	–	–	–	–	–	–

[1] "+" denotes +1 and "–" denotes –1

and

$$\text{VAR}(R3S) = \frac{1}{2}[\text{V}\hat{\text{A}}\text{R}(R3H) + \text{VAR}(R3C)]$$

where

$R'_{\text{ENC},\alpha}$ = the αth replicate ratio estimate of health characteristic x for the ENC Region with parent sample post-stratification factors used,

$R'_{\text{SW},\alpha}$ = the αth replicate ratio estimate of health characteristic x for the SW Region with parent sample post-stratification factors applied.

Variance Estimators for Estimator 2

The variances for the estimates $\hat{R}_2$ were estimated by the correct replication method only.

For $\hat{R}_2$, the variance estimators are VAR($R4H$), VAR($R4C$), and VAR($R4S$). This scheme consists of estimating the variances for the regional estimates $\hat{R}_{\text{ENC},2}$ and $\hat{R}_{\text{SW},2}$ and then combining them for a variance estimate of $\hat{R}_2$. This process is appropriate since, for the parent sample estimate, poststratification was at the regional level. For the half-sample and the complement half-sample estimates, the regional poststratification factors were calculated each time. The orthogonal patterns in tables IV and V were used in producing the variances for $\hat{R}_{\text{ENC},2}$ and $\hat{R}_{\text{SW},2}$. The formulas are

$$\hat{R}_2 = \frac{y_{\text{ENC}}\hat{R}_{\text{ENC},2} + y_{\text{SW}}\hat{R}_{\text{SW},2}}{y}$$

where

$\hat{R}_{\text{ENC},2}$ = the ratio estimate for the ENC Region with poststratification at a regional level,

$\hat{R}_{\text{SW},2}$ = the SW Region ratio estimate with poststratification at the regional level,

$$\text{VAR}(\hat{R}_2) = \frac{y_{\text{ENC}}^2[\text{V}\hat{\text{A}}\text{R}(\hat{R}_{\text{ENC},2})] + y_{\text{SW}}^2[\text{V}\hat{\text{A}}\text{R}(\hat{R}_{\text{SW},2})]}{y^2}$$

$$\text{VAR}(\hat{R}_{\text{ENC},2}:H) = \frac{1}{12}\sum_{\alpha=1}^{12}(R'_{\text{ENC},\alpha} - \hat{R}_{\text{ENC},2})^2,$$

$$\mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:C) = \frac{1}{12}\sum_{\alpha=1}^{12}(R^{*}_{\mathrm{ENC},\alpha} - \hat{R}_{\mathrm{ENC},2})^2$$

$$\mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:S) = \frac{1}{2}[\mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:H) + \mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:C)]$$

$$\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:H) = \frac{1}{12}\sum_{\alpha=1}^{12}(R'_{\mathrm{SW},\alpha} - \hat{R}_{\mathrm{SW},2})^2$$

$$\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:C) = \frac{1}{12}\sum_{\alpha=1}^{12}(R^{*}_{\mathrm{SW},\alpha} - \hat{R}_{\mathrm{SW},2})^2$$

$$\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:S) = \frac{1}{2}[\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:H) + \mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:C)]$$

$$\mathrm{VAR}(R4H) = \left\{y^2_{\mathrm{ENC}}[\mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:H)] + y^2_{\mathrm{SW}}[\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:H)]\right\}\frac{1}{y^2}$$

$$\mathrm{VAR}(R4C) = \left\{y^2_{\mathrm{ENC}}[\mathrm{VAR}(\hat{R}_{\mathrm{ENC},2}:C)] + y^2_{\mathrm{SW}}[\mathrm{VAR}(\hat{R}_{\mathrm{SW},2}:C)]\right\}\frac{1}{y^2}$$

$$\mathrm{VAR}(R4S) = \frac{1}{2}[\mathrm{VAR}(R4H) + \mathrm{VAR}(R4C)]$$

where

$R'_{\mathrm{ENC},\alpha}$ = the αth half-sample ratio estimate for the ENC Region adjusted to region level,

$R'_{\mathrm{SW},\alpha}$ = the αth half-sample ratio estimate for the SW Region adjusted to region level.

Approximate Variance Estimators

Investigators would like to obtain estimates of variance by simple methods. The results of two approximate methods thought to be solutions to this problem are discussed in this report.[g] One of the methods, VAR($R2H$), has been used both by the NCHS and the University of Michigan SRC. The discussion is based on the results of the third replication form.

For this evaluation, bias, variance, and mean square error were computed. The definitions of these measures are the same as given in the Results section. Inspection of table VI shows the biases for VAR($R3S$) for every variable decreasing with increasing sample size. The VAR($R2S$) method displays a different pattern. For family income and restricted activity days, the biases of VAR($R2S$) reduce as the sample size increases; for the other three variables, the values increase, going from design I to design II, but then drop to their lowest in design III. VAR($R2S$) does sometimes underestimate the variance of $\hat{R}_1$. This occurs in design I for the variable proportion seeing a physician and in designs II and III for the variable restricted activity days. Regardless that the biases for three variables increase, going from design I to design II, the $R2S$ method behaves better than the $R3S$ approximation. It always has the lowest bias.

Both the quantities variance and mean square error in tables VII and VIII indicate that the $R2S$ method outperforms VAR($R3S$) without any question. This procedure always has the smaller value of variance or mean square error just as in the base of bias.

A possible explanation for why VAR($R3S$) is not a good approximate variance estimator is offered. VAR($R3S$) is a variance estimator for

[g]Two more approximations of the correct variance estimates were also investigated. These results are given in Bean.[30]

Table VI. The bias of the approximate variance estimators of the sample estimator $\hat{R}_1$

Variance estimator	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R2S$)	2.469×10^4	2.164×10^{-1}	3.677×10^{-3}	1.621×10^{-3}	-1.078×10^{-5}
VAR($R3S$)	1.222×10^5	5.885×10^{-1}	3.498×10^{-2}	3.647×10^{-3}	6.536×10^{-4}
	Design II				
VAR($R2S$)	1.219×10^4	-2.705×10^{-2}	3.984×10^{-3}	1.656×10^{-3}	2.027×10^{-5}
VAR($R3S$)	5.893×10^4	1.334×10^{-1}	1.884×10^{-2}	2.489×10^{-3}	3.344×10^{-4}
	Design III				
VAR($R2S$)	4.173×10^3	-2.912×10^{-3}	1.097×10^{-3}	3.593×10^{-4}	6.765×10^{-6}
VAR($R3S$)	2.461×10^4	6.099×10^{-2}	8.263×10^{-3}	7.021×10^{-4}	1.437×10^{-4}

Table VII. The variance of the approximate variance estimators of the sample estimator $\hat{R}_1$

Variance estimator	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R2S$)	1.594×10^9	0.190×10	2.989×10^{-3}	5.298×10^{-4}	1.541×10^{-8}
VAR($R3S$)	2.678×10^{10}	0.244×10	5.020×10^{-3}	5.904×10^{-4}	9.937×10^{-7}
	Design II				
VAR($R2S$)	4.800×10^8	5.320×10^{-1}	9.071×10^{-4}	8.409×10^{-5}	3.729×10^{-9}
VAR($R3S$)	4.633×10^9	5.901×10^{-1}	1.418×10^{-3}	8.606×10^{-5}	1.810×10^{-7}
	Design III				
VAR($R2S$)	1.364×10^8	1.495×10^{-1}	8.956×10^{-5}	1.373×10^{-5}	1.109×10^{-9}
VAR($R3S$)	1.106×10^9	1.594×10^{-1}	1.807×10^{-4}	1.447×10^{-5}	4.402×10^{-8}

$\hat{R}_1$, the parent sample estimator, obtained from a weighted average of variance estimators of $\hat{R}_{ENC,1}$ and $\hat{R}_{SW,1}$. The estimating equation poststratifies the estimators to universe levels and, thus, these subuniverse estimators are adjusted to an overall distribution of ethnic-sex-age, rather than to subuniverse distribution. Questions that should be asked are What is the magnitude of the bias in $\hat{R}_{ENC,1}$ and $\hat{R}_{SW,1}$? and If the bias is not negligible, what is the effect on the variance estimator? These questions cannot be directly answered here because the biases of the subuniverse estimators were not calculated. However, VAR($R4S$) is the appropriate weighted variance estimator for a ratio estimator poststratified to regional levels, and its performance can be studied.

The analysis was based on comparison of bias, variance, and mean square error. The formulas for these measures follow:

$$\text{bias}\,[\mathrm{VAR}(R4S)] = \overline{\mathrm{VAR}(R4S)} - s^2_{\hat{R}_2}$$

$$\mathrm{VAR}[\mathrm{VAR}(R4S)] = \frac{1}{900}\sum_{j=1}^{900}\left\{[\mathrm{VAR}(R4S)]_j - \overline{\mathrm{VAR}(R4S)}\right\}^2$$

$$\mathrm{MSE}[\mathrm{VAR}(R4S)] = \mathrm{VAR}[\mathrm{VAR}(R4S)] + \text{bias}^2\,[\mathrm{VAR}(R4S)]$$

where

$$\overline{\mathrm{VAR}(R4S)} = \frac{1}{900}\sum_{j=1}^{900}[\mathrm{VAR}(R4S)]_j.$$

The biases, variances, and mean square errors for each variable and design are given in table IX. For design I, the $R4S$ method has a range of 4.813×10^{-5} for proportion seeing a physician to 9.182×10^{-3} for family income. In designs II and III, the smallest value is again for proportion seeing a physician and the largest bias for family income. Unlike the appropriate variance estimator VAR($R1S$) for the ratio estimator $\hat{R}_1$, the biases here decrease with increasing sample size, which is a desirable property.

The range for the variance of the variance estimator $R4S$ is 1.191×10^9 for family income to 2.077×10^{-8} for proportion seeing a physician in design I; in design II, the range is 3.297×10^8 to 3.748×10^{-9}; and in design III, it is 1.008×10^8 for the variable family income to 1.072×10^{-9} for proportion seeing a physician. The mean square errors exhibit a similar pattern.

Another consideration is the distribution of the sample estimate minus its expected value divided by its estimated standard error. The ratio computed was

$$\frac{R_2 - E(\hat{R}_2)}{\sqrt{\mathrm{VAR}(R4S)}}$$

where

$$E(\hat{R}_2) = \bar{\hat{R}}_2 = \frac{1}{900}\sum_{j=1}^{900}\hat{R}_{2j}.$$

The proportion of times these ratios fell within stated limits of the normal distribution are displayed in table X. Only 12 times out of a possible 60 did the empirical proportions for VAR($R4S$) surpass the normal values. The magnitudes of VAR($R4S$)'s deviations are comparable to the amounts for VAR($R1S$) with no discernible tendency for the differences to become smaller across the designs. The total error of the variance method is satisfactory, but there

Table VIII. The mean square error of the approximate variance estimators of the sample estimator $\hat{R}_1$

Variance estimator	Variable				
	Family income	Restricted activity days	Physician visits	Hospital days	Proportion seeing a physician
	Design I				
VAR($R2S$)	2.204×10^9	0.195×10	3.003×10^{-3}	5.324×10^{-4}	1.552×10^{-8}
VAR($R3S$)	3.763×10^{10}	0.278×10	6.245×10^{-3}	6.037×10^{-4}	1.421×10^{-6}
	Design II				
VAR($R2S$)	6.289×10^8	5.328×10^{-1}	9.229×10^{-4}	8.684×10^{-5}	4.140×10^{-9}
VAR($R3S$)	8.110×10^9	6.079×10^{-1}	1.774×10^{-3}	9.226×10^{-5}	2.930×10^{-7}
	Design III				
VAR($R2S$)	1.539×10^8	1.495×10^{-1}	9.320×10^{-5}	1.386×10^{-5}	1.155×10^{-9}
VAR($R3S$)	1.713×10^9	1.631×10^{-1}	2.490×10^{-4}	1.496×10^{-5}	6.470×10^{-8}

Table IX. The bias, variance, and mean square error of $R4S$ variance estimator of R_2

Variable	Bias	Variance	Mean square error
	Design I		
Family income	9.182×10^{3}	1.191×10^{9}	1.257×10^{9}
Restricted activity days	3.944×10^{-1}	0.209×10	0.224×10
Physician visits	4.517×10^{-3}	3.173×10^{-3}	3.193×10^{-3}
Hospital days	3.398×10^{-3}	1.961×10^{-3}	1.973×10^{-3}
Proportion seeing a physician	4.813×10^{-5}	2.077×10^{-8}	2.308×10^{-8}
	Design II		
Family income	2.425×10^{3}	3.297×10^{8}	3.356×10^{8}
Restricted activity days	4.057×10^{-2}	5.433×10^{-1}	5.450×10^{-1}
Physician visits	3.626×10^{-3}	9.455×10^{-4}	9.587×10^{-4}
Hospital days	2.759×10^{-3}	2.095×10^{-4}	2.171×10^{-4}
Proportion seeing a physician	2.476×10^{-5}	3.748×10^{-9}	4.362×10^{-9}
	Design III		
Family income	-8.598×10^{2}	1.008×10^{8}	1.015×10^{8}
Restricted activity days	-1.357×10^{-2}	1.405×10^{-1}	1.407×10^{-1}
Physician visits	1.214×10^{-3}	8.227×10^{-5}	8.374×10^{-5}
Hospital days	3.161×10^{-4}	1.594×10^{-5}	1.604×10^{-5}
Proportion seeing a physician	5.098×10^{-6}	1.072×10^{-9}	1.098×10^{-9}

are indications that the size of the bias is a problem. Although it would be difficult to state any general rules about drawing inferences when the stated confidence level could be in error in either direction, constructing such intervals is conceivable.

The method VAR($R2S$) can offer a savings in computer time, since the process of calculating new poststratification weights within each half-sample is eliminated. Because the comparable measures used show this method to be a good approximation, it is also compared with VAR ($R1S$). The bias of VAR($R2S$) is less in the three designs for restricted activity days and in designs I and II for the variable hospital days. In the case of variance, VAR($R2S$) has the lower value for variables restricted activity days, hospital days, and proportion seeing a physician in design I and also for hospital days in design II. The mean square errors of VAR($R2S$) are lower for the same designs and variables as the variances.

For a better understanding of the comparison between the two methods, the relative biases within a design are averaged over the variables. The relative bias is defined as

$$\text{relative bias } [\text{VAR}(R1S)] = \frac{\text{VAR}(R1S) - s^2_{\hat{R}_1}}{s^2_{\hat{R}_1}}$$

$$\text{relative bias } [\text{VAR}(R2S)] = \frac{\text{VAR}(R2S) - s^2_{\hat{R}_1}}{s^2_{\hat{R}_1}}$$

The average relative biases for VAR($R1S$) for the three designs are 0.0782, 0.0656, and 0.0363 as compared to 0.0862, 0.1231, and 0.3150, the averages for VAR($R2S$). Thus, these averages show that as the sample size increases, the magnitude of the bias in VAR($R2S$) increases, whereas for VAR($R1S$) it decreases.

To complete this discussion, the proportion of times the ratio of the sample estimate minus its expected value divided by the $R2S$ estimated variance is within certain limits is in table XI. The observation made is that 16 out of 60 empirical proportions are larger than the same proportions for the normal distribution. Con-

Table X. Proportion of times the sample estimate minus its expected value divided by *R4S* estimate of standard error is within the stated limits

Variable	Limits									
	±1.000	−1.000, 0	0, 1.000	±1.645	−1.645, 0	0, 1.645	±1.960	−1.960, 0	0, 1.960	±2.576
	Design I									
Family income . .	0.6878	0.3478	0.3400	0.8967	0.4533	0.4433	0.9422	0.4756	0.4667	0.9844
Restricted activity days . . .	0.6733	0.3300	0.3433	0.8967	0.4456	0.4511	0.9444	0.4767	0.4678	0.9789
Physician visits .	0.6578	0.3144	0.3433	0.8722	0.4289	0.4433	0.9356	0.4700	0.4656	0.9800
Hospital days . .	0.6722	0.3211	0.3511	0.8478	0.4222	0.4256	0.9033	0.4644	0.4389	0.9556
Proportion seeing a physician . .	0.6911	0.3522	0.3389	0.9011	0.4456	0.4556	0.9478	0.4678	0.4800	0.9811
	Design II									
Family income .	0.6922	0.3489	0.3433	0.8833	0.4333	0.4500	0.9344	0.4633	0.4711	0.9778
Restricted activity days . . .	0.6556	0.3044	0.3511	0.8856	0.4256	0.4600	0.9344	0.4567	0.4778	0.9778
Physician visits .	0.6878	0.3400	0.3478	0.8967	0.4578	0.4389	0.9422	0.4833	0.4589	0.9800
Hospital days . .	0.6789	0.3189	0.3600	0.8833	0.4356	0.4478	0.9367	0.4711	0.4656	0.9756
Proportion seeing a physician . .	0.7011	0.3644	0.3367	0.9167	0.4678	0.4489	0.9544	0.4867	0.4678	0.9856
	Design III									
Family income .	0.6689	0.3211	0.3478	0.8667	0.4167	0.4500	0.9233	0.4456	0.4778	0.9800
Restricted activity days . . .	0.6578	0.3356	0.3222	0.8700	0.4456	0.4244	0.9256	0.4733	0.4522	0.9744
Physician visits .	0.6867	0.3344	0.3522	0.9033	0.4467	0.4567	0.9444	0.4689	0.4756	0.9844
Hospital days . .	0.6856	0.3367	0.3489	0.8811	0.4456	0.4356	0.9278	0.4744	0.4533	0.9667
Proportion seeing a physician . .	0.6822	0.3378	0.3444	0.8856	0.4378	0.4478	0.9322	0.4589	0.4733	0.9867

trastingly, the VAR($R1S$) estimates exceeded the normal proportions only six times. This comparison is done because, ideally, when an analyst uses such an approximation, he would like to know in which direction he could be making an error. A somewhat unexpected observation is that even though the differences are not in the same direction, the average deviations are smaller for VAR($R2S$). These average differences for each statistic across the five limits and three designs are +.0145, −.0197, −.0081, −.0015, and −.0097. The average differences of VAR($R1S$) are −.0168, −.0200, and −.134. This body of empirical evidence supports the proposal of using VAR($R2S$) as an approximation.

Table XI. Proportion of times the sample estimate minus its expected value divided by *R2S* estimate of standard error is within the stated limits

Variable	Limits									
	±1.000	−1.000, 0	0, 1.000	±1.645	−1.645, 0	0, 1.645	±1.960	−1,960, 0	0, 1.960	±2.576
	Design I									
Family income	0.7433	0.3811	0.3622	0.9122	0.4644	0.4478	0.9544	0.4822	0.4722	0.9933
Restricted activity days	0.6689	0.3289	0.3400	0.8844	0.4367	0.4478	0.9333	0.4656	0.4678	0.9767
Physician visits	0.6567	0.3100	0.3467	0.8744	0.4333	0.4411	0.9244	0.4644	0.4600	0.9756
Hospital days	0.6689	0.3022	0.3667	0.8644	0.4200	0.4444	0.9200	0.4556	0.4644	0.9589
Proportion seeing a physician	0.6489	0.3400	0.3089	0.8544	0.4311	0.4233	0.9256	0.4667	0.4589	0.9833
	Design II									
Family income	0.7256	0.3711	0.3544	0.9133	0.4611	0.4522	0.9500	0.4800	0.4700	0.9844
Restricted activity days	0.6433	0.3000	0.3433	0.8878	0.4411	0.4467	0.9333	0.4611	0.4722	0.9733
Physician visits	0.6811	0.3311	0.3500	0.9011	0.4522	0.4489	0.9378	0.4744	0.4633	0.9853
Hospital days	0.6722	0.3267	0.3456	0.8800	0.4389	0.4411	0.9344	0.4767	0.4578	0.9756
Proportion seeing a physician	0.6900	0.3533	0.3367	0.9144	0.4622	0.4522	0.9522	0.4800	0.4722	0.9833
	Design III									
Family income	0.7033	0.3378	0.3656	0.9011	0.4344	0.4667	0.9489	0.4611	0.4878	0.9900
Restricted activity days	0.6533	0.3344	0.3189	0.8744	0.4478	0.4267	0.9289	0.4733	0.4556	0.9767
Physician visits	0.7011	0.3411	0.3600	0.9056	0.4433	0.4622	0.9489	0.4700	0.4789	0.9833
Hospital days	0.6933	0.3500	0.3433	0.8889	0.4533	0.4356	0.9322	0.4789	0.4533	0.9689
Proportion seeing a physician	0.7011	0.3556	0.3456	0.8822	0.4400	0.4422	0.9300	0.4589	0.4711	0.9844

☆ U. S. GOVERNMENT PRINTING OFFICE : 1975 584-528/37

VITAL AND HEALTH STATISTICS PUBLICATION SERIES

Formerly Public Health Service Publication No. 1000

Series 1. Programs and collection procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. Data evaluation and methods research.—Studies of new statistical methodology including: experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. Analytical studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and committee reports.—Final reports of major committees concerned with vital and health statistics, and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data from the Health Interview Survey.—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. Data from the Health Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian, noninstitutional population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data from the Institutional Population Surveys.—Statistics relating to the health characteristics of persons in institutions, and their medical, nursing, and personal care received, based on national samples of establishments providing these services and samples of the residents or patients.

Series 13. Data from the Hospital Discharge Survey.—Statistics relating to discharged patients in short-stay hospitals, based on a sample of patient records in a national sample of hospitals.

Series 14. Data on health resources: manpower and facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on mortality.—Various statistics on mortality other than as included in regular annual or monthly reports—special analyses by cause of death, age, and other demographic variables, also geographic and time series analyses.

Series 21. Data on natality, marriage, and divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports—special analyses by demographic variables, also geographic and time series analyses, studies of fertility.

Series 22. Data from the National Natality and Mortality Surveys.—Statistics on characteristics of births and deaths not available from the vital records, based on sample surveys stemming from these records, including such topics as mortality by socioeconomic class, hospital experience in the last year of life, medical care during pregnancy, health insurance coverage, etc.

For a list of titles of reports published in these series, write to:

Office of Information
National Center for Health Statistics
Public Health Service, HRA
Rockville, Md. 20852

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Public Health Service

HEALTH SERVICES AND MENTAL HEALTH ADMINISTRATION
5600 Fishers Lane
Rockville, Md. 20852

OFFICIAL BUSINESS
Penalty for Private Use, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF H.E.W.

HEW 396

THIRD CLASS
BLK. RATE