ALTERNATIVE VARIANCE ESTIMATION METHODS FOR THE NHIS

Mary Gessley Nixon, J. Michael Brick, Graham Kalton, Hyunshik Lee, Westat, Inc.
Mary Gessley Nixon, Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Replication, jackknife, balanced repeated replication, degrees of freedom.

1. Introduction

The National Health Interview Survey (NHIS) is a large-scale household survey of the civilian noninstitutionalized population of the United States conducted by the National Center for Health Statistics (NCHS). It began in 1957 and has undergone modifications to its content and sampling design approximately every 10 to 15 years. It consists of a core questionnaire and several specialized supplemental questionnaires. The Immunization Supplement to the NHIS, which gathers vaccination data about sample persons under the age of six, was added in 1992. In 1994, a record check component was added to the Immunization Supplement for children between the ages of 19 and 35 months. This record check component is known as the National Immunization Provider Record Check Study (NIPRCS).

The NHIS and NIPRCS provide estimates of important characteristics about the health of the United States population. Since both are samples, reliable measures of precision for those estimates are needed. The NHIS, and hence the NIPRCS, employ a complex sample design that includes stratification and several stages of sampling. Sampling weights are adjusted for nonresponse and poststratified to known population totals. The complex sample design and estimation procedures need to be taken into account in variance estimation. A complication here is that confidentiality concerns have resulted in the omission of complete stratum and PSU identifiers from public release files. Because of these concerns, NCHS offers simplified approximations to the NHIS design for use in variance estimation.

The documentation accompanying the NHIS public release file discusses two approximations to the NHIS design, termed Method 1 and Method 2. Method 1 treats the NHIS sample as a two PSU per stratum design with 187 strata and 374 PSUs. Method 2 treats the design as one with 4079 PSUs. Both methods allow the user to specify stratum and PSU easily for each record on the public release file; the user can then proceed to use any software appropriate for variance estimation from a complex design. Replication methods of variance estimation can be readily applied with Method 1 but not Method 2, which is more suited to the Taylor Series linearization method. With both methods, special care must be taken with subsamples of the NHIS, such as the NIPRCS. A discussion of subsetted data analysis is included in the variance estimation documentation included with the public use files.

This study examines an alternative to Methods 1 and 2, termed Method 3, that is designed for use with replication methods. The statistical and practical aspects of the three methods are compared using data from the 1995 NHIS.

2. Overview of the Sample Design and Estimation

The NHIS employs a multi-stage sample design, with stratification at each stage. The universe is partitioned into approximately 1900 Primary Sampling Units (PSUs), with many large PSUs included in the sample with certainty (self-representing or SR PSUs). The remaining PSUs are stratified by geography and PSU characteristics within state. In each of the non-self-representing (NSR) strata two PSUs are selected with probabilities proportional to their 1990 populations. Secondary sampling units (SSUs) are formed within each selected PSU and stratified in terms of black and Hispanic population concentration. SSUs are sampled at different rates, and within sampled SSUs households may be subsampled. Persons within sampled households may be subsampled as well.

The NHIS sampling weights account for the various stages of sampling and are adjusted for nonresponse and poststratified to age/sex/race-ethnicity totals. Additional weight adjustments were made to the 1995 data to compensate for the three week government shutdown at the end of that year.

The NIPRCS is based on a subsample of the NHIS. Children under the age of six are sampled for the Immunization Supplement. NIPRCS includes only children between the ages of 19 and 35 months (the age changed to 12 to 35 months beginning in mid-1997). All age-eligible children in a household are included, even if the child is not a sample person for the NHIS household. The final weights for the NIPRCS are based on the NHIS household weights and are adjusted for nonresponse to the Immunization Supplement.

3. Variance Estimation for the NHIS

The two methods of estimating variances discussed in the NCHS documentation were developed to enable users to compute variances using replication or linearization approaches without having extensive knowledge of the NHIS sample design and estimation procedures. Originally developed for an earlier NHIS design, they were adapted to the current design. The rationale and procedures are given by Parsons and Casady (1987) and Parsons, Chan, and Curtin (1990). These two methods are described first, then the alternative method is presented.

Method 1 is designed essentially for software that requires exactly two PSUs in each stratum. The sample is partitioned into 187 variance strata, each of which contains exactly two pseudo-PSUs. The documentation indicates that the variance strata and pseudo-PSUs are based on the actual sample design as much as possible, particularly for the NSR strata where two PSUs are sampled per stratum (collapsing is required when only one PSU was sampled in a stratum). The largest SR PSUs are treated as strata and all the SSUs within are identified as being in one of two pseudo-PSUs. Some of the smaller SR PSUs are paired into variance strata.

If BRR is used with Method 1, then a fully balanced design requires 188 replicates and replicate weights. If a stratified paired jackknife replication method is used, 187 replicates and replicate weights are needed.

For comparison with the other methods, the latter replication method is employed in the analysis reported below.

Method 2 is the alternative NCHS suggests for use with linearization software. This method treats NSR strata in almost the same way as Method 1, with two PSUs for each NSR strata and collapsing as necessary for strata in which only one PSU is sampled per stratum. The main difference relates to SR PSUs. In Method 2, the SR PSUs are partitioned into substrata based on race/ethnicity, with the substrata used as variance strata. Each SSU in a SR unit is treated as a separate PSU rather than pairing them into pseudo-PSUs as is done in Method 1.

Method 2 is considered to be more statistically efficient than Method 1 because it incorporates the design information in the NHIS for the SR PSUs. The method results in 4079 PSUs in the SR PSUs and 259 PSUs in the NSR strata. Although Method 2 should produce more stable variance estimates than Method 1, it is not well suited for use with replication methods because it would require thousands of replicates and replicate weights. NCHS documentation also warns users that, although Method 2 is more stable than Method 1, the number of degrees of freedom (df) of variance estimates computed using Method 2 are not estimated correctly using the common shortcut that takes the difference between the number of PSUs and the number of strata. The df issue for this and other methods is discussed below.

Method 3 is an alternative that takes greater advantage of the design information in the SR PSUs without requiring a large number of replicates. The main objective in designing the alternative is to produce stable and reliable variance estimates for national and domain estimates while reducing the number of replicates. The general approach is to use Method 1 design information for the NSR strata, Method 2 design information for the SR PSUs, and to take combinations of the strata to reduce the number of replicates. Combining strata for jackknife replication is the equivalent of partial balancing in BRR. Some authors refer to this as grouping strata. The strategy for combining units can take into account the important analysis domains and the stability of the variance estimates, as discussed below.

An important decision in developing the alternative under Method 3 is to determine the total number of replicates to be constructed. One consideration is that the estimates of the precision of the resultant variance estimates will be used in constructing confidence intervals and carrying out statistical tests. As the number of df of the variance estimate decreases, the ratio of the t-statistic to the z-statistic increases and the inferences are less powerful. As the number of strata that are combined increases, both the number of replicates and the df of the variance estimator decrease. Rust (1986) discusses this relationship and suggests methods to maximize the df for a specified number of replicates. If strata are combined so that only 30 effective df remain, 95 percent confidence intervals will be about 4 percent greater those based on the normal distribution. With 50 df, the 95 percent confidence interval is only 2 percent greater. Our objective was to combine strata in a way that retained at least 30 to 50 df for variance estimates for national and important domain estimates.

Distributions from the NHIS can be used to assess how many replicates are needed to give the required df. For general purpose use, the optimal procedure for combining strata is to form combined strata of equal size. The first columns of Table 1 show the number of Method 2 PSUs and the estimated population totals for each region and type of PSU (SR or NSR). Since NSR strata are less metropolitan (an important domain) and on average larger in total population than individual SR PSUs, it is best to limit the amount of combining of these strata. The maximum numbers of NSR variance strata in regions 1, 2, and 4 are 14, 38, and 18, respectively, achieved by not combining any of these NSR strata. This maximizes the number of df from the NSR strata and results in 70 replicates. If the NSR strata in region 3 are combined with NSR strata in other regions, the number of replicates remains at 70. Such a procedure also maximizes the df for region 3 variance estimates. An alternative approach is to create a total of 128 replicates, in which the NSR strata in region 3 are not combined with other NSR strata. This approach maximizes the df from the NSR strata across all the regions, but it is not examined further here.

With the number of replicates or variance strata set at 70 and the combining strategy for the NSR strata specified, the next step to is deal with the SR PSUs. Since the SSUs in the NHIS are sampled systematically using race/ethnicity concentration, a standard approach is to pair consecutive SSUs and treat these as paired selections from a stratum within the SR PSUs. The public use file does not provide the data needed to recreate the original sampling order, so the pairing is used to approximate it. This results in a large number of variance strata. These strata are then combined into 70 to correspond to the number of replicates used for the NSR strata. To increase the df for regional estimates, combining is done in variance strata in a region not assigned to the NSR strata. For example, in region 1 the NSR units are assigned to variance strata 1 to 14 and the SR units are assigned to variance strata 15 to 70. As shown in Table 1, this results in approximately a proportional allocation of the population to each variance stratum within a region.

The one exception is region 3 where the NSR units are assigned to variance strata 1-58 and this accounts for 83 percent of the 70 variance strata but only 48 percent of the population in the region. The variance estimates are more efficient if the population in the variance strata is closer to proportional, so some of the SR units were combined with the NSR units in this region. Figure 1 depicts the overall assignment of the units to the 70 variance strata for Method 3.

Some of the advantages expected from the Method 3 approach to forming variance strata and replicates are that the variances should be more consistent than Method 1 because of improved handling of SR PSUs, and the df for domains such as region, metropolitan status, and race/ethnicity should be reasonably large.

4. Properties of the Variance Estimators

The three variance estimators can be evaluated in terms of their consistency, unbiasedness, and precision. Under most conditions, all three methods are consistent. Rao and Shao (1996) and Valliant (1996) show that combining a large number of SSUs in SR PSUs as is done in Method 1 can result in an inconsistent variance estimator. However, because the number of strata is large and the contribution to the variance from each SR PSU is relatively small in the NHIS, the results of Rao and Shao and Valliant may not apply here and the Method 1 variance estimator should be consistent. All three are unbiased, as discussed in Wolter (1985) and Rust (1986).

The main criterion remaining for evaluating the three variance estimators is precision. Since the full jackknife variance estimator reproduces the textbook variance estimate for linear statistics, we use this as the benchmark for comparing the three methods. Following the development in Rust (1986), the variance of the full jackknife variance estimator is

$$V\left(v_{FJ}(\hat{\theta})\right)=\sum_{h=1}^{H}\frac{W_h^4\sigma_h^4}{n_h^2}\left[\frac{\beta_h-3}{n_h}-\frac{2}{n_h-1}\right] \qquad (1)$$

where $\hat{\theta}$ is the sample estimate, $W_h = N_h/N$, and σ_h^2 and β_h are the population variance and kurtosis in stratum h for the cluster estimates.

Combining strata decreases the precision of the estimator. The variance of combined strata jackknife estimator is

$$V(v_{CJ}(\hat{\theta}))=\sum_{h=1}^{H}\frac{W_h^4\sigma_h^4(\beta_h-3)}{n_h^3}+2\sum_{g=1}^{G}\left(\frac{\sum_{h\in g}^{H_g}W_h^2\sigma_h^2/n_h}{(l_g-1)}\right)^2 \qquad (2)$$

where g denotes a combined stratum consisting of H_g original strata and l_g is the number of PSUs in the combined stratum.

These expressions can used to evaluate the precision of a variance estimator if the values of W_h, σ_h^2, and β_h are known, but generally this is not the case and consistent sample estimates are substituted. Kish (1965) shows β tends to 3 for large clusters in sampling human populations. We assume $\beta_h = 3$. We use the sample estimate pq/n_h multiplied by a design effect for σ_h^2. The design effect is an average computed separately for SR and NSR PSUs; p was assumed to be 0.5. The estimated population proportion in each stratum h was used for W_h. We explain below how expressions (1) and (2) are applied to the three methods of variance estimation proposed for the NHIS below.

For Method 2, equation (1), the full jackknife variance estimator, can be applied directly. Since Method 2 assumes that the SSUs are selected by SRS within the SR PSUs while systematic sampling is actually used, an alternative (Method 2a) that assumes stratified sampling of the SSUs (with the same stratification used in Method 3) is used in this evaluation to be more comparable to the other methods.

In Method 1, the exact method of combining and collapsing strata cannot be determined from the public use information. As a result, expression (2) cannot be evaluated. An alternative, called Method la, uses expression (2) and assumes stratified sampling within the SR PSUs for the comparisons. The Method 3 stratification is used and strata are combined into the Method 1 variance strata.

For Method 3, stratified sampling within the SR PSUs and combining of strata are essential features, and expression (2) is used to estimate the variance of the variance estimator. The combining of strata is as defined in Figure 1.

To relate these quantities to the df of the variance estimate, the Satterthwaite approximation of the number of df, $r = 2V(\hat{\theta})^2 \div V(v(\hat{\theta}))$, is used. Since the numerator is constant for all three methods of variance estimation, the ratios of the variances of the variance estimators measure the relative precision of the three methods.

5. Comparison of the Methods

As expected, all three methods provide adequate degrees of freedom for national estimates. As shown in Table 2, Method 1a uses 187 replicates and gives 72 df; Method 2a gives 501 df, but requires 2,167 replicates; Method 3 give 66 df for national estimates, but requires only 70 replicates. For a 95 percent confidence interval, the appropriate t-value for Method 2 is 1.97, and for Methods 1a and 3 it is 1.99, a very small difference. The gain in precision for Method 2a is not great, despite the large increase in df.

Note that for Methods 1a and 2a, the number of df is far smaller than the commonly used 'rule of thumb', which estimates the df by the number of PSUs minus the number of strata. On the other hand, with Method 3, the 'rule of thumb' provides a reasonable approximation. The 'rule of thumb' works well with Method 3 because the combined strata are formed to be of roughly equal size, as needed for the rule to apply.

Table 2 also presents the estimated number of df for the three methods for important domain estimates including region, race/ethnicity, poverty status, and MSA status. The goal was to have 30 to 50 df available for the domain estimates. Domains were chosen to include characteristics such as region with uneven distribution over strata, and those more evenly distributed, such as poverty. As expected, Method 2a provides the largest number of df for all domains. However, Method 3 provides adequate df for all the domains (the lowest being 42 df for the non-MSA domain) and requires far fewer replicates.

Overall, Method 3 provides more df than Method 1a for most of the domain estimates, with many fewer replicates. Method 1a has more df for the Non-MSA estimate, but has only 7 df for Region 1 and 10 df for the Hispanic domain. For domains with uneven distribution across strata, such as region and race characteristics, Method 3 is clearly an improvement over Method 1a. If 128 replicates been used for Method 3, so as not to combine any NSR strata, the df for the non-MSA would probably have been higher. This evaluation shows that Method 3 is clearly preferred to Method 1a for computational and statistical efficiency.

Table 2. Estimated number of degrees of freedom for Methods 1a, 2a, and 3

Domain	Method 1a	Method 2a	Method 3
National	72	501	66
Region 1	7	169	61
Region 2	24	122	57
Region 3	51	139	59
Region 4	12	113	56
Poverty	53	154	52
Black	26	164	55
Hispanic	10	53	32
MSA	48	482	63
Non-MSA	60	60	42
Number of replicates	187	2,167	70

6. Summary

Replication variance methods provide an attractive way to allow users to estimate sampling error appropriately from a public release data file, while eliminating the need to attach potentially sensitive design information to the file. The data provider can use all relevant sampling information to create replicate weights for the file, including adjustments for nonresponse and poststratification. (The effect of these nonresponse and poststratification adjustments could not be evaluated in this study because the needed replicate weights are not provided in the public release file.) A disadvantage of the replication method is that it can be computationally intensive if the number of replicates is high. Combining variance strata is one way to reduce the number of replicates, and Method 3 appears to be a reasonable strategy for doing this for the NHIS.

Method 3 was used to produce variance estimates for characteristics of children estimated for the NIPRCS. Despite the fact that the NIPRCS is a relatively small subsample of the full NHIS sample, no special precautions were needed to deal with the subset. The replicate weights from the NHIS were adjusted for nonresponse in the NIPRCS using the same procedures employed for the full sample weights. The resulting replicate weights contained all the design information required for computing variances.

7. References

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

Parsons, V., and Casady, R. (1987). Variance estimation and the redesigned National Health Interview Survey. *Proceedings of Survey Research Methods Section of the American Statistical Association,* 406-411.

Parsons, V., Chan, J., and Curtin, L.R. (1990). Analytic limitations to current national health surveys. *Proceedings of Survey Research Methods Section of the American Statistical Association,* 736-741.

Rao, J.N.K., and Shao, J. (1996). On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association*, 91, 343-348.

Rust, K. (1986). Efficient replicated variance estimation. *Proceedings of Survey Research Methods Section of the American Statistical Association*, 81-87.

Valliant, R. (1996). Limitations of balanced half-sampling. *Journal of Official Statistics*, 12, 225-240.

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

Table 1. Method 3 assignment of 70 variance strata for the NHIS

Region	Type	Number of PSUs	Population percent	Region percent	Number of variance strata	Percent in variance strata	Assigned variance strata
	NSR	29	3.5%	17.9%	14	20.0%	01 to 14
1	SR	994	16.1%	82.1%	56	80.0%	15 to 70
	NSR	76	10.2%	42.9%	38	54.3%	15 to 52
2	SR	766	13.6%	57.1%	32	45.7%	01 to 14/53 to 70
	NSR	117	16.7%	47.5%	58	82.9%	01 to 58
3	SR	1210	18.5%	52.5%	70	*	59 to 70.01 to 59
	NSR	37	5.2%	24.3%	18	25.7%	53 to 70
4	SR	1109	16.1%	75.7%	52	74.3%	01 to 52
	NSR	259	35.6%				
Total	SR	4079	64.4%				
		4338	100.0%				

* In this category, the SSUs are first assigned to variance strata 59 to 70 until they are roughly equal in size to the NSR variance strata, then they are distributed across all the variance strata.

Figure 1. Assignment of variance strata by region and type

Variance Strata	Region 1		Region 2		Region 3		Region 4	
	NSR	SR	NSR	SR	NSR	SR	NSR	SR
1	✓			✓	✓	✓		✓
2	✓			✓	✓	✓		✓
3	✓			✓	✓	✓		✓
4	✓			✓	✓	✓		✓
5	✓			✓	✓	✓		✓
6	✓			✓	✓	✓		✓
7	✓			✓	✓	✓		✓
8	✓			✓	✓	✓		✓
9	✓			✓	✓	✓		✓
10	✓			✓	✓	✓		✓
11	✓			✓	✓	✓		✓
12	✓			✓	✓	✓		✓
13	✓			✓	✓	✓		✓

Figure 1. Assignment of variance strata by region and type (continued)

	Region 1		Region 2		Region 3		Region 4	
	NSR	SR	NSR	SR	NSR	SR	NSR	SR
14	✓			✓	✓	✓		✓
15		✓	✓		✓	✓		✓
16		✓	✓		✓	✓		✓
17		✓	✓		✓	✓		✓
18		✓	✓		✓	✓		✓
19		✓	✓		✓	✓		✓
20		✓	✓		✓	✓		✓
21		✓	✓		✓	✓		✓
22		✓	✓		✓	✓		✓
23		✓	✓		✓	✓		✓
24		✓	✓		✓	✓		✓
25		✓	✓		✓	✓		✓
26		✓	✓		✓	✓		✓
27		✓	✓		✓	✓		✓
28		✓	✓		✓	✓		✓
29		✓	✓		✓	✓		✓
30		✓	✓		✓	✓		✓
31		✓	✓		✓	✓		✓
32		✓	✓		✓	✓		✓
33		✓	✓		✓	✓		✓
34		✓	✓		✓	✓		✓
35		✓	✓		✓	✓		✓
36		✓	✓		✓	✓		✓
37		✓	✓		✓	✓		✓
38		✓	✓		✓	✓		✓
39		✓	✓		✓	✓		✓
40		✓	✓		✓	✓		✓
41		✓	✓		✓	✓		✓
42		✓	✓		✓	✓		✓
43		✓	✓		✓	✓		✓
44		✓	✓		✓	✓		✓
45		✓	✓		✓	✓		✓
46		✓	✓		✓	✓		✓
47		✓	✓		✓	✓		✓
48		✓	✓		✓	✓		✓
49		✓	✓		✓	✓		✓
50		✓	✓		✓	✓		✓
51		✓	✓		✓	✓		✓
52		✓	✓		✓	✓		✓
53		✓		✓	✓	✓	✓	
54		✓		✓	✓	✓	✓	
55		✓		✓	✓	✓	✓	
56		✓		✓	✓	✓	✓	
57		✓		✓	✓	✓	✓	
58		✓		✓	✓	✓	✓	
59		✓		✓		✓	✓	
60		✓		✓		✓	✓	
61		✓		✓		✓	✓	
62		✓		✓		✓	✓	
63		✓		✓		✓	✓	
64		✓		✓		✓	✓	
65		✓		✓		✓	✓	
66		✓		✓		✓	✓	
67		✓		✓		✓	✓	
68		✓		✓		✓	✓	
69		✓		✓		✓	✓	
70		✓		✓		✓	✓	