ISSUES RELATING TO THE USE OF JACKKNIFE METHODS IN THE NATIONAL IMMUNIZATION SURVEY

Philip J. Smith, CDC; K.P. Srinath, Michael P. Battaglia, Abt Associates; Barry I. Graubard, NCI; Lawrence Barker, CDC; David C. Hoaglin, Martin Frankel, Abt Associates; Meena Khare, NCHS
Philip J. Smith, CDC/NIP, Corporate Square Blvd., Bldg. 12, Room 2109, Atlanta, GA 30333

Key Words: Bias, Complex sampling design, Nonsampling weight adjustment, Variance estimation.

1. Introduction

The analysis of survey data requires the application of special statistical methods to adjust for the effects of a complex sampling design that has been implemented to satisfy analytic objectives, contain survey costs, and ease the administrative burden of carrying out the survey. The complex sampling design may entail stratification, clustering, unequal probability sampling, and the use of auxiliary sampling frames. Särndal, Swensson, and Wretman [1] provide a recent overview of statistical methods that account for the complex sampling design's effect on the expected error that is incurred in estimating finite population parameters. It is for this reason that in this paper we refer to the statistical adjustments that are made to survey weights to account for the effects of a complex sampling design as *sampling design adjustments.*

Carefully designed and implemented estimation methodologies for sample surveys account for *more* than the effects of the complex sampling design on errors in estimation, however. Additional adjustments are routinely made to survey weights with the objectives of either reducing the potential bias of estimated population parameters or to accept increased potential bias in return for reducing their expected mean squared error. Among the adjustments that are intended for these purposes are

- adjustments for unit nonresponse,
- poststratification to known population totals,
- weight trimming, and raking,
- adjustments for noncoverage of the target population by the sampling frame, and
- adjustments for expected differences in response patterns between sampled units that provide information on key survey topics compared to units that do not provide information.

Because these adjustments are not associated with the error incurred in estimating population parameters as a result of the complex sampling design, we refer to these adjustments as *nonsampling weighting adjustments* to distinguish them from the *sampling design adjustments*. Although nonsampling weighting adjustments are made to reduce bias of point estimates, if the statistical methods used to estimate standard errors do not account for those adjustments, estimates of standard errors may be biased. This may have an untoward effect on the coverage properties of calculated confidence intervals.

For example, a common and simple method that is used in sample surveys for estimating the standard error of a complicated statistic has been described by Woodruff [2]. This method has come to be known as the Taylor series linearization or delta method and is implemented in commercial software that is widely used to analyze complex survey data. As noted by Rust and Rao [3], it is conceptually straightforward to account for some of the nonsampling error adjustments explicitly in the Taylor series linearization method. However, these authors indicate that, as the number and complexity of nonsampling weighting adjustments increase, the difficulty in accounting for these adjustments using the Taylor series method increases. The extent to which this difficulty increases makes this approach to variance estimation inexpedient or impossible, when a full accounting for nonsampling weighting adjustments in the estimation process is desired. As a result, one would expect that estimated standard errors computed using the Taylor series method would be biased because these methods do not routinely account for all of the features of both the complex sampling design and the nonsampling weight adjustments.

As an alternative to the Taylor series method for estimating standard errors from complex surveys, replication methods may be used. Rust and Rao [3] provide a recent review of these methods. These methods offer a simple way of accounting for all of the features of both the complex sampling design and the nonsampling weight adjustments.

Also, Rust and Rao summarize the literature in which replication methods and the Taylor series method for estimating standard errors are compared with respect to their bias, coverage rates, and asymptotic properties. This literature was pioneered by Kish and Frankel [4], followed by work by Krewski and Rao [5], Lemeshow and Levy [6], Kovar et al. [7], Valliant [8], Rao and Wu [9], Shao and Wu [10], and others. Results from these investigations indicate that jackknife replication methods give similar results to those obtained by using Taylor series methods, both yielding estimates of standard errors with negligible bias. However, we note that in all of this research, the implicit assumption was that nonsampling weighting adjustments were not required: There was no comparison of methods in the usual situation in which numerous and complex nonsampling weight adjustments are required. In this case, the conclusions obtained in this previous research may not generalize to the more common situation in which the Taylor series approach does not account for nonsampling weight adjustments, but the replication methods do.

This paper presents a case study of the bias that is incurred by estimated standard errors using the Taylor series method that does not account for the numerous nonsampling weight adjustments in the National Immunization Survey (NIS). To make this evaluation, we compare results to those obtained using the jackknife replication method described by Särndal et al. [1, p. 415]. In making comparisons, we tailor the jackknife to account for all of the features of the complex sampling design and all of the nonsampling weight adjustments used in the NIS.

In Section 2 of this paper we give a synopsis of the complex sampling design of the NIS. Also, we summarize how the NIS sampling weights are constructed, highlighting sampling design and nonsampling weight adjustments. In Section 3 we review the Taylor series method for obtaining estimated standard errors of estimated vaccination coverage rates and give an example of how the method yields biased estimates when it does not account for nonsampling weight adjustments. Also, we describe how a properly constructed jackknife estimate of variance accounts for the complex sampling design and all of the nonsampling weight adjustments. We review the literature that shows that a properly constructed jackknife estimate of variance has negligible bias.

In Section 4 we present the results of our investigation, showing the extent of the bias in standard errors using Taylor series methods for the NIS. In Section 5 we conclude with a discussion of the results and related literature.

2. The NIS Sampling Design

The NIS covers 78 Immunization Action Plan (IAP) areas, which include the 50 states and 28 urban areas, including the District of Columbia. Each IAP area represents a stratum of the sampling design within which the NIS samples independently. Within each IAP area, the design of the NIS includes 2 phases of sampling: a random-digit-dialing (RDD) survey of households followed by a mail survey of vaccination providers of eligible children in sampled households.

In the RDD sampling phase of the NIS, the respondent in a sampled household with eligible children is asked to provide information on the demographic and socio-economic characteristics of the household. Also, in the RDD interview the respondent is asked for consent to contact eligible children's immunization providers to obtain information about their vaccination histories. If verbal consent is obtained, the vaccination providers are mailed a questionnaire from which vaccination histories are obtained. Ezzati-Rice et al. [11] and Zell et al. [12] give a more detailed description about the sampling design of the NIS.

2.1 Sampling Design and Nonsampling Weight Adjustments

In the NIS, 9 separate weight adjustments are made. Among these, only 2 are sampling design adjustments, and the remaining 7 are nonsampling error adjustments. The following paragraphs describe each of these adjustments, the order in which they are made, and their purpose, as well as indicating whether the adjustment is a sampling design weight adjustment (SDA) or a nonsampling weight adjustment (NSA).

The base sampling weight *(SDA).* Each child with data in the NIS receives a base sampling weight, equal to the reciprocal of the probability of selecting the household's telephone number into the sample. Specifically, this weight is the ratio of two totals for that IAP area: (1) the number of telephone numbers in the population, and (2) the number of telephone numbers drawn from the population.

Base sampling weight trimming *(NSA).* In an RDD sample it is possible for a household to be sampled from an IAP area but to physically be located in an adjacent IAP area. Because a large range in the base weights can substantially increase the variance of estimates, each child's base weight is not allowed to exceed three times the base weight for the IAP area in which the child resides, as calculated above.

Multiple residential telephones *(SDA).* A household with two or more residential telephone numbers has a proportionally higher probability of being selected into the RDD sample. To preserve the relationship between the base sampling weight and this probability, an adjustment is made for the number of non-business voice-use telephone numbers reported in the household.

Multiple residential telephone adjustment weight trimming *(NSA).* Division of a household's trimmed base sampling weight by the number of non-business voice-use telephone numbers reported in the household can introduce considerable variation in the adjusted weights. This is known to increase sampling variability. Thus, the reported number of these telephones is limited to no more than 3. In doing this we accept bias in order to reduce variance.

Multi-level household nonresponse *(NSA).* Unit nonresponse can occur at several points in the NIS interviewing process. At each point a different amount of information is available about the nonresponding telephone number. To reduce potential bias, the NIS applies a separate weighting-class adjustment for each of three amounts of information:

1) The interviewer has identified an eligible household, but the interview has not been completed;
2) The survey has reached a household, but nothing more is known; and
3) It is unknown whether the telephone number is residential.

Within each of a set of cells or classes the adjustment increases each respondent's base sampling weight to account for the nonrespondents. For example, where each nonrespondent is known to be an eligible household, each respondent's base weight is multiplied by the ratio of the number of respondents and nonrespondents to the number of respondents. The cells are defined by IAP area, the residential directory-listed status of the sample telephone number, and telephone-exchange-level demographic and socioeconomic characteristics.

Nontelephone coverage *(NSA).* Random-digit dialing yields a sample of children in households that have telephones, but the NIS aims to measure vaccination coverage levels for all children 19 to 35 months of age. Data from the National Health Interview Survey (NHIS) indicate that vaccination levels are generally lower among nontelephone children than among telephone children. In some IAP areas a substantial proportion of age-eligible children reside in nontelephone households. To compensate for such potential noncoverage bias, the NIS employs a weight adjustment procedure described by Battaglia et al. [13].

Poststratification *(NSA).* Poststratification separates the actual sample into cells defined by characteristics that are related to noncoverage and vaccination status. Then the weighted distribution of completed interviews over the cells is brought into agreement with a corresponding set of population totals. The purpose of this adjustment is to reduce bias incurred by obtaining samples that have weighted totals that do not agree with known population totals of variables that are believed to be associated with vaccination coverage. In RDD surveys these differences often arise from differential nonresponse.

Provider-reported vaccination history nonresponse *(NSA).* The weighting estimation methodology that is currently in use for the NIS has been designed specifically to also adjust vaccination coverage estimates for provider-reported vaccination history nonresponse bias (Smith et al. [14, 15]). Within each IAP area, the methods achieve this by grouping sampled children into adjustment cells according to the similarity of their response propensities to have a provider-reported vaccination history. As a first step in forming adjustment cells, a response propensity model was developed using logistic regression. The response propensity is the probability that a sampled child has a provider-reported vaccination history.

The provider immunization history nonresponse adjustment cells were formed by sorting the children in an IAP area by their response propensities and then forming five cells with equal unweighted counts of sample children. To adjust for vaccination history nonresponse bias, within each adjustment cell, children with vaccination histories are assigned a revised set of weights that are obtained by dividing their 1st-phase sampling weights by the cell-specific weighted response rate. By dividing the 1st-phase sampling weights of children who have provider vaccination information by their adjustment-cell-specific weighted response rate, these children more fairly represent all of the children in the cell as a whole.

Raking *(NSA).* The revised weights may not match poststratification totals used to construct 1st-phase sampling weights. Also, the revised weights may not match the 1st phase sample weighted totals of other variables that are known to be important predictors of being up-to-date. To reduce bias attributable to these differences and to maintain the nonresponse bias adjustment, we rake the revised weights to match poststratification totals, outcome predictor totals, and the adjustment-cell-specific 1st-phase sampling weight totals.

3. The Taylor Series and Jackknife Methods for Estimating Standard Errors

Wolter [16] provides a complete description of commonly used methods for estimating standard errors of estimated finite population parameters. In Section 3.1 we review the Taylor series linearization method applied to the problem of estimating the variance of a ratio estimate for a subdomain of the target population when the sampling design corresponds to a stratified 1-stage cluster sampling design, as in the NIS. In Section

3.2 we describe how the jackknife can be applied to estimating the variance of this ratio estimate, accounting for nonsampling weighting adjustments, also.

For the purposes of this section, let

N_h = the number of primary sampling units (PSUs, or households in the case of the NIS) in stratum h;

n_h = the number of PSUs sampled in stratum h, h=1, … , L;

M_{hi} = the number of subjects in PSU i of stratum h belonging to the target population;

m_{hi} = the number of subjects in PSU i in stratum h who were sampled in the survey;

W_{hij} = the overall sampling weight for subject j sampled in PSU i of stratum h, accounting for all sampling weight and nonsampling weight adjustments;

Y_{hij} = 0 when subject j in PSU i of stratum h is not up-to-date on a specific vaccination, and Y_{hij} =1 when the child is up-to-date; and

$\delta_{hij} = 1$ when subject j in PSU i of stratum h belongs to the domain of interest, and $\delta_{hij} = 0$, otherwise.

Letting $Y_h = \sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} \delta_{hij} Y_{hij}$ and $T_h = \sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} \delta_{hij}$, the true but unknown vaccination coverage rate is

$$\theta = \frac{\sum_{h=1}^{L} Y_h}{\sum_{h=1}^{L} T_h}.$$

3.1 Taylor Series Linearization Variance Estimation

Let $\hat{Y}_h = \sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} \delta_{hij} W_{hij} Y_{hij}$ and $\hat{T}_h = \sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} \delta_{hij} W_{hij}$.

Then the combined ratio estimator (Cochran [17, p. 165]) of the vaccination coverage rate for the domain of interest is

$$\hat{\theta} = \frac{\sum_{h=1}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{T}_h}. \quad (1)$$

Letting $Z_{hij} = \frac{\delta_{hij} W_{hij} (Y_{hij} - \hat{\theta})}{\hat{T}_h}$ denote the linearized value of (1) and letting $Z_{hi} = \sum_{j=1}^{m_{hi}} Z_{hij}$ and $\overline{Z}_h = \frac{\sum_{i=1}^{n_h} Z_{hi}}{n_h}$, the Taylor series estimate of the variance of $\hat{\theta}$ is

$$\hat{V}_T(\hat{\theta}) = \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (Z_{hi} - \overline{Z}_h)^2 . \quad (2)$$

3.2 Jackknife Replication Variance Estimation

Wolter [16, p. 175] indicates that, although the jackknife is conceptually simple, when applying this method to stratified sampling designs, special care must be exercised to obtain a correct expression for the jackknife variance. For a stratified 1-stage cluster design used in the NIS, let the sampled PSUs that yielded a completed interview in stratum h be partitioned at random into A_h groups of approximately equal numbers of PSUs (all sample telephone numbers that did not yield a completed interview are also randomly partitioned into equal sized groups). After omitting the *a*th group from stratum h, let $B_{(ha)}$ denote the remaining complement of PSUs in stratum h. We refer to $B_{(ha)}$ as the *a*th replicate sample in stratum h.

For sampled units belonging to the replicate sample $B_{(ha)}$, we compute a separate set of replicate sampling weights, $\{W_{hij}^{(ha)}\}$. These replicate weights are constructed as if *only* the *a*th replicate sample, $B_{(ha)}$, had been obtained in the sample in stratum h. In this case, all of the sampling design and nonsampling weight adjustments described in Section 2.1 are applied to construct $\{W_{hij}^{(ha)}\}$. Also, $W_{hij}^{(ha)} = 0$ for all sampled units in stratum h not belonging to $B_{(ha)}$. Note that, in general, $W_{hij}^{(ha)} \neq W_{hij}$. In this respect differences between $W_{hij}^{(ha)}$ and W_{hij} reflect variability that can be attributed to sampling design *and* nonsampling weight adjustments.

Let $\hat{Y}_h^{(a)} = \sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} \delta_{hij} W_{hij}^{(ha)} Y_{hij}$ and $\hat{T}_h^{(a)} = \sum_{i=1}^{n_n}\sum_{j=1}^{m_{hi}} \delta_{hij} W_{hij}^{(ha)}$. Then the replicate estimator of θ based on what remains in the entire sample after omitting the ath group from stratum h is

$$\hat{\theta}_{(ha)} = \frac{\hat{Y}_h^{(a)} + \sum_{h' \neq h}^{L} \hat{Y}_{h'}}{\hat{T}_h^{(a)} + \sum_{h' \neq h}^{L} \hat{T}_{h'}}.$$

The jackknife estimator of the variance of (1) is

$$\hat{V}_J(\hat{\theta}) = \sum_{h=1}^{L} \frac{(A_h - 1)}{A_h} \sum_{a=1}^{n_h} (\hat{\theta}_{(ha)} - \hat{\theta})^2. \qquad (3)$$

Note that differences between $\hat{\theta}_{(ha)}$ and $\hat{\theta}$ in (3) reflect variability from differences between the replicate weights $W_{hij}^{(ha)}$ and the overall weights W_{hij}. Because differences between $W_{hij}^{(ha)}$ and W_{hij} reflect variability that can be attributed to sampling design *and* nonsampling weight adjustments, (3) adjusts the estimated variance of $\hat{\theta}$ to reflect variability from both of these sources. Wolter [16] shows that, to second-order moments, the jackknife estimator of variance is unbiased. This result is congruent with the empirical findings of Kish and Frankel [4] and other research summarized by Rust and Rao [3].

4. Results

To evaluate the bias of standard errors obtained using the Taylor series method, the jackknife method was used with $A_h = 40$ in each stratum of the NIS. This value was selected to provide a sufficiently large number of degrees of freedom for variance estimation that would be suitable for many different types of statistical analysis that may be conducted with data from a single stratum.

The jackknife estimate $\sqrt{\hat{V}_J}$ and Taylor series estimate $\sqrt{\hat{V}_T}$ of the standard error of the estimated 4:3:1 vaccination series coverage rate (4 or more doses of the diphtheria-tetanus-pertussis (DTP) vaccine, 3 or more doses of polio vaccine, and 1 or more doses of the measles-mumps-rubella vaccine (MMR)) were obtained for each state using data from the 1999 NIS. Using these statistics for each state, we computed

- $\sqrt{\hat{V}_T / \hat{V}_J}$, an estimate of the bias of the Taylor series estimated standard error, relative to the estimated standard error obtained using the jackknife estimate,
- $W_{95\%,T} / W_{95\%,J} = z_{.975}\sqrt{\hat{V}_T} / t_{.975,\phi}\sqrt{\hat{V}_J}$, the relative half-widths of the 95% confidence intervals formed using jackknife and Taylor series estimates of the standard errors, and
- $1 - \alpha_T = 2 \times T_\phi(z_{.975}\sqrt{\hat{V}_T / \hat{V}_J}) - 1$, the estimated actual confidence coefficients of the nominal 95% confidence intervals formed using the Taylor series estimate of the standard error, assuming that the jackknife estimate of the standard error provides the bias-corrected estimate of the standard error of the estimated 4:3:1 vaccination coverage rate.

Here $t_{.975,\phi}$ is the upper 97.5%-ile of Student's t distribution with ϕ degrees of freedom, and $z_{.975}$ is the upper 97.5%-ile of the standard normal distribution. For each state, the degrees of freedom ϕ equals $A_h - 1$ times the number of strata (IAP areas) contained within the state. Also, $T_\phi(\bullet)$ denotes the cumulative distribution function of Student's t distribution with ϕ degrees of freedom.

Table 1 gives selected quantiles of the distribution of the ratio of these statistics estimated from each of the 50 states and the District of Columbia. This table shows that the range of $\sqrt{\hat{V}_T / \hat{V}_J}$ is from 0.796 to 1.352, suggesting that estimated Taylor series standard errors may be biased by 20% or more. Also, Table 1 shows the 75%-ile of $\sqrt{\hat{V}_T / \hat{V}_J}$ is 1.080, suggesting that the Taylor series method for estimating standard errors tends to underestimate standard errors of 4:3:1 state-level estimated vaccination coverage levels.
Also Table 1 shows that the statistic $W_{95\%,T} / W_{95\%,J}$ ranges from 0.771 to 1.331, indicating that the width of the 95% confidence interval computed using a Taylor series estimate of the standard error may be as much as 23% too narrow or 33% too wide. We note that only 3% of the difference in widths of the confidence intervals is accounted for as a result of the relative difference between $t_{.975,\phi}$ and $z_{.975}$.

Finally, Table 1 shows that the distribution of the actual confidence coefficient, $1-\alpha_T$, ranges from 0.873 to 0.990, indicating that the actual confidence coefficient of 95% confidence intervals can differ greatly from the nominal 95%. Among the 50 states and the District of Columbia, 40 have actual confidence coefficients that differ from the nominal level by 1 percentage point or more.

References

[1] Särndal, C.E., Swensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer-Verlag.

[2] Woodruff, R.S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association.* 66: 411-414.

[3] Rust, K.F. and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5: 283-310.

[4] Kish, L. and Frankel, M. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B,* 36: 1-22.

[5] Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics,* 9: 1010-1019.

[6] Lemeshow, S. and Levy, P.S. (1979). Estimating the variance of ratio estimates in complex sample surveys with two primary units per stratum – a comparison of balanced repeated replication and jackknife techniques. *Journal of Statistical Computing and Simulations*, 8: 191-205.

[7] Kovar, J.G., Rao, J.N.K., and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statistics,* 16: 25-46.

[8] Valliant, R. (1990). Comparisons of variance estimators in stratified random and systematic sampling. *Journal of Official Statistics,* 6: 115-131.

[9] Rao, J.N.K. and Wu, C.F.J. (1985). Inference from stratified samples: second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association,* 80: 620-630.

[10] Shao, J. and Wu, C.F.J. (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *The Annals of Statistics,* 20: 1571-1593.

[11] Ezzati-Rice, T.M., Zell, E.R., Battaglia, M.P., Ching, P.L.Y.H. and Wright, R.A. (1995). The Design of the National Immunization Survey. Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 668-672.

[12] Zell, E.R., Ezzati-Rice, T.M., Battaglia, M.P., and Wright, R.A. (2000). National Immunization Survey: The Methodology of a Vaccination Surveillance System. *Public Health Reports*, 115(1), 65-77.

[13] Battaglia, M. P., Malec, D.J., Spencer, B.D., Hoaglin, D.C., and Sedransk, J. (1995). Adjusting for Noncoverage of Nontelephone Households in the National Immunization Survey. Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 678-683.

[14] Smith, P.J., Rao, J.N.K., Daniels, D., Battaglia, M.P., Ezzati-Rice, T., Khare, M. (2000). Compensating for Nonresponse Bias in the National Immunization Survey Using Response Propensities. To appear in the Proceedings of the Section on Survey Research Methods of the American Statistical Association.

[15] Smith, P.J., Rao, J.N.K., Hoaglin, D., Daniels, D., Battaglia, M.P. (2000). Alternative Methods for Compensating Provider-Verified Vaccination History Nonresponse Bias in the National Immunization Survey. To appear in the Proceedings of the Section on Survey Research Methods of the American Statistical Association.

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

[17] Cochran, W.G. (1979). *Sampling Techniques.* John Wiley and Sons: New York.

[18] Rubin, D.B., and Schenker, N. (1987), Interval Estimation from Multiply-Imputed Data: A Case Study Using Census Agriculture Industry Codes, *Journal of Official Statistics*, 3, 375-387.

[19] Rubin, D.B., and Schenker, N. (1986), Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse, *Journal of the American Statistical Association*, 81, 366-374.

Table 1: Selected quantiles of the distribution of the ratio of estimated standard errors, 95% confidence interval widths, and actual confidence coefficients of nominal 95% confidence intervals.

Percentile	$\sqrt{\hat{V}_T / \hat{V}_J}$	$W_{95\%,T} / W_{95\%,J}$	$1-\alpha_T$
0%	0.796	0.771	0.873
25%	0.942	0.921	0.930
50%	0.976	0.954	0.939
75%	1.080	1.058	0.961
100%	1.352	1.331	0.990