ISSUES RELATED TO ANALYSIS OF A SINGLE YEAR OF DATA FROM THE ANNUAL NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEYS

Margaret D. Carroll, National Center for Health Statistics; Jenny Chia, Orkland Corp.; and Jill Montaquila, Westat
Margaret D. Carroll, National Center for Health Statistics, 6525 Belcrest Road Room 900, Hyattsville, Maryland 20782-2003

KEY WORDS: Variance estimation, jackknife, balanced repeated replication, replication, linearization, degrees of freedom

1. Introduction

This paper deals with one of the issues that needs to be considered in the analysis of annual data from the first year of the current National Health and Nutrition Examination Survey (NHANES). It is one of a number of methodological investigations needed to determine whether analysis of annual NHANES surveys is feasible. In particular, variance estimation methods appropriate for analysis of annual NHANES data are considered and software packages that can be used for estimation and analysis are discussed.

Section 2 provides some background on the NHANES survey program and the current NHANES in particular. The issue of variance estimation for annual NHANES data is discussed in Section 3. In Section 4, some software packages available for analysis of data from complex sample surveys such as NHANES are considered. Finally, Section 5 contains a discussion and summary.

2. Background

The National Center for Health Statistics / Centers for Disease Control (NCHS/CDC) began conducting periodic national health and nutrition examination surveys (NHANES) in 1970. In 1999 NCHS/CDC changed the NHANES into a series of ongoing annual health and nutrition examination surveys. These surveys are designed to provide information on the health and nutritional status of the United States civilian non-institutionalized population. A unique feature of these surveys is the collection of health data by means of medical examinations carried out for a nationally representative sample of the U.S. population. Some key characteristics of the NHANES surveys are shown in Table 1.

Data collection for NHANES is done in three components: A household screener, an interview, and a medical examination. The primary objective of the screener is to determine whether any household members are eligible for the interview and examination. The interview collects household-, family-, and person-level data on health and nutrition characteristics. The examination includes anthropometric measurements, tests such as eye and dental examinations, a dietary component and the collection of blood and urine specimens for laboratory testing.

Like previous NHANES surveys, the current NHANES used a multistage national area probability sample. The stages of selection were 1) primary sampling units (PSUs); 2) segments within PSUs; 3) households within segments; and 4) sampled persons within households.

To standardize their administration, the examinations are carried out in mobile examination centers (MECs). Considering the time and the cost involved in moving a MEC between survey locations, the sample size per PSU must be large enough to produce an efficient workload at each PSU. In addition, to reduce the amount of travel necessary for respondents to visit a MEC, and thereby increase the likelihood of achieving high response rates, the PSUs for NHANES are typically defined as individual counties. In a few cases, adjacent counties were combined to keep PSUs above a certain minimum size. Segments comprised Census blocks or groups of blocks.

For this analysis each sampled person who participated in the current NHANES (respondent) was assigned a sample weight which may be thought of as an indication of the number of people in the population that that individual represents. The sample weight is the reciprocal of the probability of selection, with an adjustment for non-response and a poststratification adjustment that aligns the

survey estimates with current external population estimates.

The differences in the sample sizes, study populations and designs for the four cycles of NHANES should be considered when comparisons are made across various NHANES surveys. For example, it should be noted that until NHANES III, the NHANES surveys did not include persons 75 years or older and that NHANES I and NHANES II did not oversample Hispanics.

3. Variance Estimation for analysis of annual NHANES surveys

In a complex sample survey setting, variance estimates will be biased if computed using standard statistical software packages that assume simple random sampling. In order to produce approximately unbiased (and consistent) estimates of variance based on data from the NHANES surveys, procedures that incorporate the sample weights and account for the complex sample design must be applied.

Unlike previous NHANES surveys (which were designed for multi-year data collection), the 1999 NHANES consisted of only 12 PSUs which were selected from two nationally representative panels of the National Health Interview Survey (NHIS) PSUs. Eleven of these PSUs were non-certainty PSUs and one was a certainty PSU. The 11 noncertainty PSUs were selected from 2 NHIS panels[1] with probability proportional to a measure of size. The certainty PSU was split to create 2 pseudo-PSUs, for a total of 13 pseudo-PSUs for variance estimation for NHANES 1999.

The 1999 NHANES annual sample is limited in its analytic capabilities. For example, although the 1999 NHANES annual sample is nationally representative, it was selected from only 12 PSUs and the sample sizes for specific race/ethnicity-sex-age subdomains are relatively small. Because of the small sample sizes for analytic subdomains, analysts should use only broad subdomains, to ensure adequate precision in the survey estimates. The small number of PSUs poses challenges for variance estimation. With a small number of PSUs, direct, design-based variance estimates will be relatively unstable. Additionally, with so few PSUs and with the design used to select the 1999 NHANES PSUs, it is not feasible to account for the effect of subsampling from the NHIS PSUs on the variances of the estimates. In this section, we first describe design-based methods of variance estimation for complex sample survey data, then give an overview of model-based alternatives aimed at producing more stable variance estimates.

Two common design-based approaches are available for directly estimating variances of estimates from complex survey data: linearization and replication.

Linearization. For the linearization approach, nonlinear estimates are approximated by linear ones for the purpose of variance estimation. The linear approximation is derived by taking the first-order Taylor series approximation for the estimator. Standard variance estimation methods for linear statistics are then used to estimate the variance of the linearized estimator.

The 1999 NHANES sample of PSUs was selected using systematic sampling from a sorted list, with no explicit stratification. Due to the lack of explicit stratification and the small number of PSUs, it is recommended, if linearization is used for variance estimation, that a linearization variance estimator for unstratified sample designs be applied. The linearization variance estimate is obtained by computing the appropriate sum of squared differences between the linearized estimate for each pseudo-PSU and the mean linearized estimate.

Replication. Replication methods provide a general means for estimating variances for the types of complex sample designs and weighting procedures usually encountered in practice. The basic idea behind the replication approach is to

- Form G replicate subsamples which resemble the entire sample;
- Estimate the statistic of interest for each replicate $\hat{\theta}_{(g)}$ using only the elements in that replicate; and

[1] PSUs in the NHIS sample were partitioned into 4 nationally representative subgroups, or "panels," 2 of which were made available to NHANES in order to facilitate sample linkage with NHIS.

- Use the variability of these G statistics to estimate the variance of the full sample estimator $\hat{\theta}$.

Two examples of replication methods are Jackknife 1 (JK1) and Balanced Repeated Replication (BRR). BRR forms replicates by deleting half of the PSUs at a time whereas JK1 deletes only one PSU at a time. By retaining practically the entire sample in each replicate the JK1 method produces more stable estimates of variance than BRR especially for small subdomains. The formula for the JK1 estimate of variance is

$$v(\hat{\theta}) = \frac{(G-1)}{G}\sum_{g=1}^{G}(\hat{\theta}_{(g)} - \hat{\theta})^2 \qquad (1)$$

See Wolter (1985) for further descriptions of both the replication and linearization approaches.

One of the main advantages of the replication approach is its ease of use at the analysis stage. The same estimation procedure is used for the total sample and for each replicate. The variance estimates are then readily computed using a simple procedure. Furthermore, the same procedure is applicable to most statistics desired, such as means, percentages, ratios, and correlations (see Efron, 1982). These estimates can also be calculated for analytic groups or subpopulations. Another important advantage of the replication approach is that it provides a way to account for adjustments that are made in weighting, such as adjustments for nonresponse and poststratification. By separately computing the weighting adjustments for each replicate, it is possible to reflect the effects of nonresponse adjustment and poststratification in the estimates of variance.

Model-based methods. As an alternative to the direct, design-based variance estimates introduced above, model-based methods may be used to obtain more stable variance estimates. Average design effects (deff) or generalized variance functions (GVFs) may be used to smooth the design-based variance estimates. Both of these approaches require the computation of design-based variance estimates.

4. Software

Two software packages that are used for the analysis of data from complex sample surveys are WesVar and SUDAAN. WesVar uses replication methods: BRR, Jackknife and Fay's method.[2] Versions of SUDAAN prior to version 7.5 provided for only linearization. Version 7.5 and later versions of this software package provide for BRR and Jackknife also, with limitations. See SUDAAN version 7.5 manual for more details.

SUDAAN release 7.5 does not incorporate an option for direct estimation of variance by means of the JK1 method. Its Jackknife design option does not utilize JK1 replicate weights in estimating variances. However, SUDAAN can employ the JK1 method indirectly (see Exhibit 1).

```
PROC DESCRIPT DATA=IN.EXAMPLE
DESIGN=BRR;
WEIGHT WTSF1M;
REPWGT WTSF1MRA WTSF1MRB ….
WTSFMRM/ADJFAY=12;
VAR BMXBMI;
SUBGROUP RIDAGE2;
LEVELS 15;
TABLE RIDAGE2;
```

Exhibit 1. SUDAAN code to obtain JK1 variance estimates

SUDAAN and WesVar produce identical variance estimates for

- Means,
- Percentages,
- Geometric means,
- Standardized means,
- Totals, and
- Regression coefficients[3].

[2] Fay's method is a variation of BRR. When forming replicates, rather than entirely dropping one of the two PSUs in each stratum (as is done in BRR), one PSU is down-weighted, and the other is weighted up to compensate. For more details about this procedure, see Westat (2000).

[3] We recommend against using regression analysis on the NHANES 1999 annual sample due to the small number of degrees of freedom available for analysis.

In SUDAAN with the BRR design option and the ADJFAY= $\frac{1}{(1-k)^2}$ option on the REPWGT statement, the formula

$$v(\hat{\theta}) = \frac{1}{G(1-k)^2}\sum_{g=1}^{G}(\hat{\theta}_{(g)} - \hat{\theta})^2 \qquad (2)$$

is used to estimate the variance.

If $\frac{1}{(1-k)^2} = (G-1)$, then (2) is equivalent to (1). For NHANES 1999, G-1=12.

Exhibit 1 contains an example of SUDAAN code that could be used to produce JK1 variance estimates for estimates from NHANES 1999.

Computation of variances of percentiles. The point estimates of percentiles obtained in WesVar and SUDAAN will be identical if the "no group" option is used in WesVar and the "ungrouped" option is used in SUDAAN.

However, in estimating variances and confidence intervals these two software packages use different methods of estimation. SUDAAN uses the standard replication formula (1) with the BRR design option and the ADJFAY option on the REPWGT statement. WesVar uses Woodruff's Method. (See Sarndal *et al.*, 1992.) Through this method the point estimate and the corresponding 95 percent confidence limits are first estimated by means of the cumulative percent distribution of the variable of interest and its inverse. The standard error of the percentile is then estimated by dividing the difference between the upper and lower 95 percent confidence limits by twice the critical value based on the t distribution with the appropriate number of degrees of freedom. These two methods yield estimates that are asymptotically equivalent; i.e., for sufficiently large sample sizes, the results are identical or nearly identical. However, the results could be substantially different especially for extreme percentiles such as the 95th percentile, if the sample size is small or if the distribution is extremely skewed.

Figures 1-5 illustrate these points. In Figures 1-3 the 95 percent confidence intervals obtained from SUDAAN and WesVar for blood lead based on data from the 1999 NHANES are compared for subdomains with increasing sample sizes. As the sample sizes increase, the 95 percent confidence limits produced by the two software packages become identical or nearly identical. Figure 4 compares results from the two software packages for Urine diethylthiolphosphate – an extremely skewed distribution (Figure 5). Even though the sample size is fairly large (n=703), the two software packages produce the 95 percent confidence limits for the uppermost percentiles that are substantially different.

Computation of design effects. The standard definition of the design effect of a statistic is the ratio of the variance of a statistic based on the complex sample design to the variance of a statistic under weighted simple random sampling (*i.e.*, sampling with replacement).

SUDAAN provides 4 different definitions for design effects using four different denominator definitions:

1. Simple random sampling, assumes a fixed total sample size but a variable subgroup sample size- the option in SUDAAN in versions prior to version 7.5
2. Simple random sampling based on weighted data assumes a fixed subgroup sample size.
3. With replacement sampling selected with unequal probabilities of selection; or
4. Model based estimates of variance.

Option 4 is the default option in SUDAAN version 7.5. SUDAAN's option 2 yields the standard definition of the design effect. This is the definition of the design effect used in WesVar.

5. Discussion

NHANES can visit only a small number of PSUs each year. Consequently, variance estimates for annual estimates are relatively unstable because they are based on only a small number of degrees of freedom.

Replication methods recommended for variance estimation for possible use in annual surveys such as 1999 NHANES can be implemented

using either WesVar or SUDAAN version 7.5 or subsequent versions. To produce more stable variance estimates, a model-based approach (such as average design effects or GVFs) may be used to smooth the design-based variance estimates.

Variance estimation is only one of a number of statistical issues that need to be addressed to determine whether analysis of data from annual NHANES surveys are feasible. Examples of other issues include disclosure avoidance analysis for annual data and stability of estimates from regression modeling approaches.

References

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Kovar M.G. and Johnson C. (1986). Design effects from the Mexican American portion of the Hispanic Health and Nutrition Examination Survey: A strategy for analysts. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.

Sarndal, C.E. Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer-Verlag.

Shah, Babubhai V, Barnwell, Beth G, Bierler, Gauyle S., SUDAAN USER'S MANUAL Release 7.5 (1997). Research Triangle Institute, Research Triangle Park, NC

Westat (2000). *WesVar 4.0 User's Guide*. Rockville, MD: Westat.

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

Table 1. Comparison of various characteristics of the NHANES surveys

Characteristic	NHANES I	NHANES II	NHANES III	NHANES 1999
Years	1971-1974	1976-1980	1988-1994	1999
Geographic coverage	United States (excluding Alaska and Hawaii)	United States (including Alaska and Hawaii)	United States (including Alaska and Hawaii)	United States (including Alaska and Hawaii)
Domains for oversampling	Low income: children aged 1-5 years; women aged 20-44 years; persons aged 65 years and over	Low income: children aged 6 months-5 years; persons aged 60-74 years	52 subdomains were predesignated consisting of age-sex groups for black, Mexican American, and other persons. Target sample sizes were established for the subdomains	53 subdomains were predesignated consisting of age-sex groups for black, Mexican American, and other persons. Target sample sizes were established for the subdomains
Age range	1 – 74 yrs.	6 mos. – 74 yrs.	2 mos. and older	No restriction on age
Number of survey locations	100	64	89	12
Sample size	28,043	27,801	39,695	5,325
Interviewed sample size	27,482	25,286	33,994	4,150
Examined sample size	20,749	20,322	30,818	3,812

Figure 1. Blood lead concentrations for sample persons aged 1-5 years (Sample size = 254)

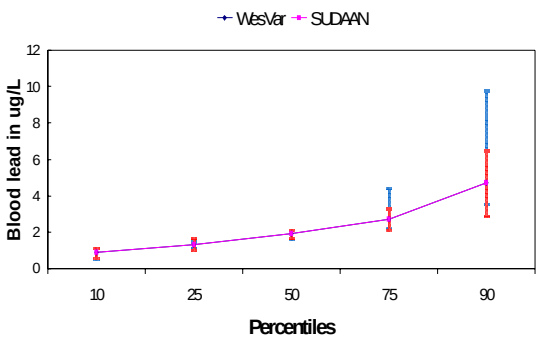

Figure 2. Blood lead concentrations for sample persons aged 40-59 years (Sample size = 471)

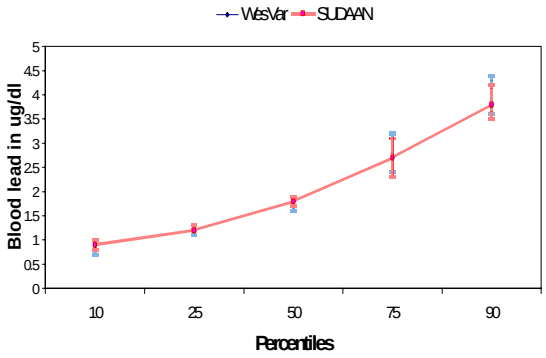

Figure 3. Blood lead concentrations for sample persons ages 1 year and older (Sample size=3,189)

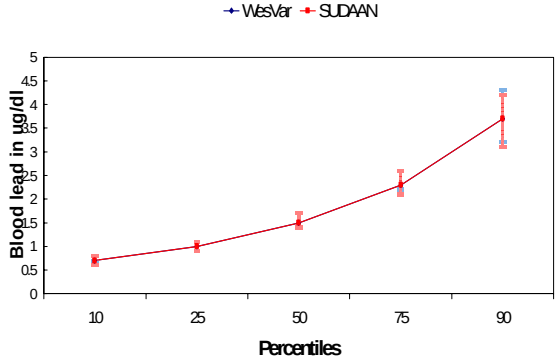

Figure 4. Urine diethylthiolphosphate concentration for SPs aged 6-59 years (Sample size = 703)

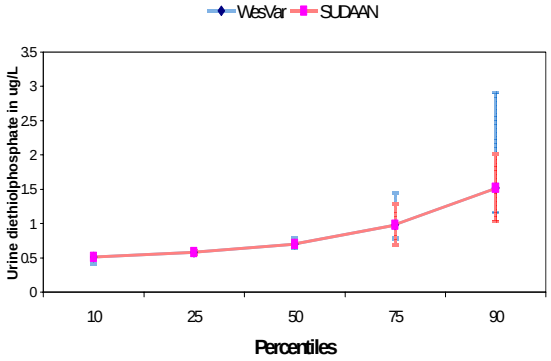

Figure 5. Distribution of Urine Diethythiolphosphate for sample persons aged 6-59 years (Sample size=703)