Estimating Variances for Health and Disability Domains in the National Health Interview Survey

Mitchell P. LaPlante, University of California
201 Filbert Street, Suite 500, San Francisco, CA 94133

KeyWords: Variances, complex surveys, health, disability

1. Introduction

The National Health Interview Survey (NHIS) is designed to produce precise (low variance) estimates of the health of the civilain noninstitutionalized United States population at low cost (Massey, Moore, Parsons, & Tadros, 1989). The NHIS uses a stratified multi-stage cluster sample design. In the NHIS, clustering reduces the precision of estimates more than the improvement in precision due to stratification. Elements within clusters are often correlated. SRS methods of estimating variances, which ignore this correlation, are likely to be imprecise, though the degree of imprecision is often unknown, particularly in studies that analyze the data in novel ways. When SRS estimates are highly imprecise, study conclusions based on statistical tests that assume SRS will be biased. Analysts of NHIS public use tapes are therefore wise to guage the extent that SRS estimates are imprecise and whether to undertake alternative methods.

This paper provides an evaluation of the accuracy of strategies for the estimation of variances of selected statistical estimates for health and disability domains derived from the National Health Interview Survey (NHIS). This evaluation is part of a study of the distribution of health insurance coverage in the U.S. and the relationship of different health coverage statuses with utilization of health care.

Direct complex variance estimation methods, which incorporate the structure of a clustered survey design, are more precise than SRS and therefore more appropriate for confidence interval construction and hypothesis testing. Three methods for direct computation of variances are Taylor series approximation, jackknife repeated replication, and resampling (bootstrap and other resampling methods). Evaluation of these different methods has indicated that no one method is consistently more accurate, and the choice of method may depend on the availability and cost of software (Lee, Forthofer, and Lorimer, 1989, p.44).

Direct complex methods are generally perceived as more cumbersome than SRS methods, adding an arduous step to the analysis process requiring more analysis time and increasing analysis costs. *Generalized complex variance estimation methods* are often used by producer statistical agencies because they may overcome the imprecision of SRS methods without requiring direct estimation of variances each time new statistics are developed from a survey. Two models are commonly employed. In the *relative variance model* (RV), the variance of a statistic is modeled as a function of the magnitude of the statistic. This approach is more popular among the statistical agencies because only the models and their parameters need be published rather than a variance estimate for each statistic. Relative variance models are typically used by producer agencies with extensive publication goals. Another model is the *average design effect model* (ADE) in which the ratio of the variance of complex to the SRS variance estimates is assumed to be constant (Kish & Frankel, 1974). SRS variance estimates can then be multiplied by an average design effect to yield more precise estimates. In both approaches, models are typically developed for classes of related statistics (e.g. narrow, medium, or wide range) based on a representative subset of domains (e.g. age, sex, education, etc.) representing the range of statistical quantities the producer agency intends to publish from a survey (Choi & Casady, 1983). Formulas with a few parameters are developed to predict variances of statistics generated from a particular survey. For variances of proportions, rates, or percents, formulas based on the magnitude of the statistic and of the denominator are employed. For variances of averages and ratio estimates, additional formulas based on the magnitude of the statistic and of the numerator and denominator are employed, although they may be crude. Generalized variance approaches are of interest mainly to analysts who do not have software available with which to directly estimate complex variances or for whom such estimation may be too tedious or expensive.

Generalized variances will likely be biased for any particular statistic (because they are averages of variances of related statistics), but they may be less subject to sampling variability than direct estimates. However, no theoretical basis appears to exist for the claim of added stability (Wolter, 1985).

Relative variance models are used frequently in national surveys. They have been used in the Current Population Survey (CPS) since 1947 (Hanson, 1978) and the NHIS since 1958. Complex variances are directly estimated for related classes of statistics (usually totals of populations or number of events) for representative domains and then modeled as a function of the magnitude of the statistic.

Generalized variance approaches have particular limitations for outside analysts. The RV parameters for the NHIS that are provided by NCHS are designed to give readers an order of magnitude estimate of the variance of published statistics and may not apply well to statistics that are different from those published by NCHS. Furthermore, the formulas are cumbersome to apply and may introduce error. When thousands of estimates are involved, the application of RV formulas and parameters becomes daunting. Also, analysts will often be interested in developing statistics for which generalized variance models (either RV or ADE) have not been developed. Complex variances need then be estimated directly and compared with SRS estimates to determine the degree of imprecision that exists.

(Cohen, 1982) has evaluated average design effect and relative variance methods in the 1977 National Medical Care Expenditure Survey (NMCES). It was found that the average design effect technique was significantly more accurate than the relative variance technique for population totals and means.

The magnitude of design effects and the accuracy of generalized variance approaches used in the NHIS are not well determined. In this study, a method of directly estimating variances using Taylor series approximation is compared with two generalized variance methods — average design effect (ADE) and relative variance (RV) (the latter based on curves estimated by NCHS) — and with conventional SRS methods for estimating variances of proportions, means, and ratio statistics from the NHIS. Quantities examined include the proportion with various types of health insurance coverage, the proportion uninsured, and mean annual physician contacts and hospital days per discharge by health insurance coverage status and health status characteristics.

2. NHIS sample design

The NHIS is a stratified multistage probability design. For every week of the year, a representative sample of the US population is drawn without replacement. The weekly samples are additive over time and yield annual, semi-annual, and quarterly statistics. Detailed characteristics of the sample design are described by Massey, Moore, Parsons, & Tadros (1989), and an overview is provided below.

Four stages of sampling are involved in the NHIS: 1) primary sampling units within strata, 2) secondary sampling units (clusters of households within selected PSUs) called SSUs, 3) households within SSUs, and in some cases, 4) persons within households for supplements.

In the design for the NHIS, the country is divided into 1,900 primary sampling units (PSUs) consisting of single counties, groups of contiguous counties, or metropolitan statistical areas (MSAs). The PSUs are grouped into strata defined by socioeconomic and demographic variables from which a sample is selected. Strata are defined so as to be demographically and socioeconomically diverse while individuals within strata will be demographically and socioeconomically similar. The largest 52 strata, called self-representing (SR), are selected with probability of one. The remainder are called non-self-representing (NSR), of which there are 73, and two PSUs are drawn from each with probability proportional to sample size. Thus, in the current design, 198 PSUs are drawn.

Within each PSU, a second-stage sample of area segments is drawn that are geographically distributed within the PSU. Segments are subdivided into clusters ranging from 4 to 9 housing units. Demographic and health information is obtained for all individuals in occupied housing units within the clusters. The NHIS uses an area rather than list sample. The NHIS is stratified by race, age, gender, and geography. The NHIS oversamples within PSUs area segments in which blacks are more highly represented.

As a precaution, to accommodate budget shortfalls, the NHIS consists of four panels, designed to efficiently create subsamples of 75, 50, or 25 percent of the annual NHIS sample size. Each of the panels is equally distributed within Census regions and is representative of the U.S. In forming panels, NSR strata are paired within regions by combining strata that are similar with respect to the original stratification variables. This results in four NSR PSUs per stratum pair which were then randomly assigned to one of the 4 panels. For estimating variances, the unpaired strata are used. For SR PSUs, the assignment to panels depends on size. The 12 largest PSUs were divided across all 4 panels, the 14 medium-sized PSUs were divided into two panels, and the 26 smallest of the SR PSUs were included in only one panel. For purposes of variance estimation, the 26 smallest PSUs were subdivided into two Pseudo-PSUs.

In the final result for the 1989 sample, a full sample, there are 73 strata with 2 NSR PSUs per, and 52 strata with 2 or 4 SR PSUs per, for a total of 125 strata. Pseudo-PSUs are defined for the purpose of variance calculations, of which there are 274. This consists of 146 NSR and 128 SR (reflecting that the SR PSUs have been grouped into panels— 12x4+14x2+26x2=128). In sum, the design includes 274 PSUs grouped into 125 strata.

3. Method of analysis and evaluation of accuracy

This analysis examines three statistical quantities: 1) percent of population with various health insurance coverage statuses (e.g., private, Medicare, Medicaid, uninsured) by health and disability characteristics, 2) mean annual physician visits per person by health insurance coverage status and health characteristics, and 3) mean length of hospital stay per hospital discharge by health insurance coverage status and health and disability characteristics. Twelve health insurance coverage categories were defined. Health is measured by self-ratings of health status (6 categories), and disability by activity limitation (4 categories), work limitation (4 categories), and need for personal assistance in basic life activities (3 categories). These variables were cross-classified resulting in 107 health and disability domains.

All statistics are estimated using weights to inflate to national totals. These weights take into account 1) the inverse of the probability of selection, 2) household nonresponse adjustment, 3) first-stage ratio adjustment, and 4) second-stage ratio adjustment (post-stratification) (Massey,Moore,Parsons, & Tadros, 1989). For the quantities described above, variances were estimated directly by Taylor series approximation. Two other generalized variance strategies — an average design effect model and a relative variance model — were also examined. These three methods are described below.

Directly estimated variances (Taylor series approximation)

In large samples, the approximation to the variance of a complex sample estimator, Y, is of the form

$$E\left\{\sum_{i=1}^{m}\frac{\partial Y}{\partial Y_i}\left[Y_i - E(Y_i)\right]^2\right\} \quad (1)$$

where the partial derivatives $\partial Y/\partial Y_i$ are evaluated at their expected values (Woodruff, 1971). This expression considers the first-order approximation of the deviation of Y from its expected value. This method of approximation is applied to the PSU totals within each stratum, and is a weighted combination of the variation across PSUs within the same stratum. Y may be a complex sample proportion, mean, or ratio of two variables.

Variance estimates for proportions with health insurance and mean physician contacts were obtained using the software SESUDAAN and for average stay per hospital discharge using the software RATIOEST, employing national inflation weights. The software requires at least one pair of PSUs per estimated statistic. Estimates with inadequate numbers of PSUs were discarded for this analysis. All duplicate categories in the various cross-nestings of health and disability variables were discarded. For mean doctor visits and mean hospital stays, two health insurance categories with low cell sizes were excluded. The number of resulting estimates for each of the three statistical quantities is shown in Table 1.

Generalized variances

Two models are considered: the average design effect (ADE) and the relative variance (RV). The average design effect estimate was calculated by multiplying the SRS estimate of variance by the mean estimated variance from the Taylor series approximation by the following steps. First, the design effect, or DEFF, is defined as

$$DEFF(d) = VAR_{CMPLX}(d) / VAR_{SRS}(d) \quad (2)$$

The average DEFF is then

$$\overline{DEFF} = \sum_{d=1}^{D} DEFF(d)/D \quad (3)$$

where D is the total number of domain estimates. Then the predicted variance is simply $\overline{DEFF}$ multiplied by the SRS sampling variance

$$VAR_{ADE}(d) = \overline{DEFF} \times VAR_{SRS}(d). \quad (4)$$

With the ADE model, SRS variances need to be estimated which are then multipled by $\overline{DEFF}$.

Relative variance estimates were computed using model parameters for relevant classes of NHIS statistics and formulas published in Current Estimates (Adams & Benson, 1990). Parameters are based on models of the form

$$RV(x) = VAR(x)/x^2 = a + b/x + e \quad (5)$$

where e is an error term. The models are estimated using a representative sample of population domains (64 in 1985) (Massey et al., 1989) for which variances have been computed using Taylor series. NCHS publishes a and b parameters for a 10 classes of estimates from the NHIS which can be employed using appropriate formulas to predict variances for a variety of statistics.

The formula used for the percent with a particular type of health coverage is

$$VAR_{RV}(p) = \frac{bp(100 - p)}{y} \quad (6)$$

where y is the weighted population value for the denominator and b=3,640. The formula used for a weighted mean or ratio (r=x/y) is

$$VAR_{RV}\left(\frac{x}{y}\right) = \left\{\frac{x}{y} \times \sqrt{\frac{VAR_{RV}(x)}{x^2} + \frac{VAR_{RV}(y)}{y^2} - 2r\frac{SE_{RV}(x)}{x}\frac{SE_{RV}(y)}{y}}\right\}^2 \quad (7)$$

referred to as Rule 4 in Current Estimates.

For mean annual physician contacts (based on a two-week reference period, the formula for the numerator (weighted number of contacts) is

$$VAR_{RV}(x) = ax^2 + bx \tag{8}$$

with a=0.0000282 and b=166,000. For the denominator, we use (8) and substituting y (weighted number of persons) for x, a=0.0000307 and b=3,640. For average hospital days per discharge, for the numerator (weighted number of days) we use (8) with a=0.00194 and b=82,300 and for the denominator we also use (8) substituting y (weighted number of discharges) for x, with a=0.000187 and b=6,220. Note that in (7) the term for the correlation (r) between numerator and denominator cannot be estimated from aggregate statistics. The quantity r is often assumed to be negligible (Cox & Cohen, 1985), but perhaps incorrectly so. If r is positive, formula (7) will overestimate the true variance, and if r is negative, formula (7) will underestimate the true variance.

Evaluation of accuracy

(Cox & Cohen, 1985) suggest the average relative absolute difference between direct and predicted estimates of variance as a measure of the accuracy of generalized variance strategies. This is defined as

$$\overline{A} = \frac{1}{D}\sum_{d=1}^{D}\frac{|\hat{s}_p^2(d) - \hat{s}_o^2(d)|}{\hat{s}_o^2(d)} \tag{9}$$

where $\hat{s}_o^2(d)$ is the direct variance estimate obtained by the Taylor series approximation and $\hat{s}_p^2(d)$ is the predicted variance for the d-th domain estimate, and D is the total number of domain estimates. As predicted estimates approach Taylor series estimates, $\overline{A}$ approaches zero. This measure incorporates deviations due to bias or instability.

Estimates of accuracy were computed by (9) for the average design effect model and the relative variance curve model. For the average design effect model

$$\hat{s}_p^2(d) = \overline{DEFF} \times VAR_{SRS}(d) \tag{10}$$

and for the relative variance model

$$\hat{s}_p^2(d) = VAR_{RV}(d). \tag{11}$$

As a check on the accuracy of using SRS variance estimates without correcting for $\overline{DEFF}$, results were obtained using the following measure

$$\hat{s}_p^2(d) = VAR_{SRS}(d). \tag{12}$$

If Taylor series estimates are closely related to SRS estimates, then the average relative absolute difference from (10) and (12) should be similar and complex estimates can be ignored. This situation may arises in surveys with low clustering or low intracluster correlations on specific variables in which case, as defined by (3), the $\overline{DEFF}$ would be close to 1.

4. Results

Approximately 30% of the estimates of the percent with insurance or mean annual physician contacts had standard errors computed by Taylor series exceeding 30% of the magnitude of the estimate, a standard measure of reliability used in NHIS publications. About 21% of the estimates of the ratio of hospital days to discharges had standard errors exceeding 30% of the magnitude of the estimate.

Correlation coefficients between SRS and the Taylor series were computed. For the percent with insurance, the correlations of Taylor series and the relative variance with SRS estimates are both very high—.97 and .98. Considering mean physician contacts, the correlation of Taylor series with SRS is also very high—.99—but the correlation coefficient falls to .83 for relative variance estimates. For the ratio of hospital days to discharges, the correlation of Taylor series with SRS remains high—.95—but the correlation coefficient falls to .78 for relative variance estimates. While these correlations show there is a closer linear relationship between the Taylor series estimates and SRS estimates than for the RV and SRS estimates, they do not measure the accuracy of these estimates.

Design effects (DEFF) for the Taylor series and the relative variance curve methods are shown in Table 2. Considering the percent with insurance, the mean DEFF was 1.37 for the Taylor series estimates and 1.71 for the relative variance curve estimates. The Taylor series estimates with the greatest design effects occurred for the largest population totals and tended to skew the distribution. Though the Taylor series estimates had a lower DEFF, there was more variation than for the RV estimates. For physician contacts the mean DEFF for the Taylor series is 1.14, while the mean DEFF for the relative variance curves is only .79, but exhibits very high variation. For average length of stay in hospital, the mean DEFF for the relative variance estimates is 13.24—over 12 times as high as the mean DEFF for Taylor series estimates, which is clearly an aberration, highly variable, and highly skewed.

The average relative absolute difference measures the accuracy of predicted variances from the average design effect model and the relative variance curve model compared to direct Taylor series estimates (Table 3). Given that the average design effects for the Taylor series estimates were found to be greater than 1 for all three statistical quantities, the first column shows, as we expect, that Taylor series estimates produce higher estimates of variance than the SRS estimates (H_0: average relative absolute difference =0). Considering the percent with insurance, despite the fact that both the Taylor series and the RV estimates were highly correlated with SRS estimates, together with the fact that the mean DEFF for the RV method was higher than that for the direct Taylor series estimates, these results show that the RV method significantly overestimates the variance. For all three statistical quantities, the average design effect model exhibits significantly greater accuracy than the relative variance model. For the percent with insurance, the RV model is 3.4 times more inaccurate than the average design effect model for mean annual physician contacts and almost two orders of magnitude more inaccurate for the ratio of hospital days per discharge.

5. Conclusion

Estimating complex variances for insurance coverage and health care utilization statistics from the NHIS directly using Taylor series linearization, it is found that the mean design effects were 1.04 for average hospital days per discharge, 1.14 for mean annual physician visits per person, and 1.37 for the percent with insurance coverage. A prediction model using average design effects multiplied by SRS estimates of variance was found to be significantly more accurate than a prediction model based on NHIS relative variance curves for all three statistical quantities examined. This suggests that variances can be better approximated by an average design effect model than by NCHS variance curves for proportions, means, and ratio estimates. When direct complex estimation of variances is not feasible, the average design effect model provides a reasonable alternative to NCHS relative variance models if the average design effect is known.

The relative variance model produces especially variable results for means and ratio estimates. This is most likely explained by the fact that the relative variance curve model does not incorporate an estimate of the correlation between numerator and denominator for means and ratios. If this correlation is positive, the relative variance curve model will overestimate the variance and if negative, will underestimate the variance.

In this analysis, only PSU clustering was evaluated. Incorporation of clustering in the sample design within secondary sampling units could produce different results. These design parameters however are not available on public use tapes.

These results suggest that SRS methods result in lower precision which may bias some analyses. The degree of imprecision depends on the type of statistic, being higher in this analysis for a proportion statistic and lowest for a ratio statistic. Analysts should not rely on published relative variance curve models to produce estimates of means and ratios as they can be much less accurate than using SRS estimates.

These results could be sensitive to the specific domains analyzed. Health status, disability, and insurance coverage domains may exhibit a different variance pattern than the domains employed by NCHS in estimating RV curves. Nevertheless, the fact that the variances based on RV curves are much closer to those based on the direct Taylor series method for a simple proportion than they are for mean and ratio statistics implicates the formulas used for computing RV estimates. It is recommended that NCHS review the policy of estimating variances for means and ratio statistics in the NHIS, such as annual physician visits or hospital days per hospitalized person. Specifically, it is suggested that NCHS consider adding tables of standard errors, estimated using direct complex methods, for mean and ratio statistics presented in Current Estimates instead of advocating the use of formulas based on relative variances of totals. It would also appear that direct estimates would be more accurate for proportions. This recommendation is not without precedent. In recent Bureau of the Census reports, i.e., a recent report on household wealth and asset ownership based on the Survey of Income and Program Participation, (Eargle & U.S. Bureau of the Census, 1990), tables of directly computed standard errors have been included for mean wealth estimates.

6. References

Adams, P. F., & Benson, V. (1990). Current Estimates from the National Health Interview Survey, 1989. Vital Health Stat, 10(176), 1-121.

Choi, J. W., & Casady, R. J. (1983) Variance in the National Health Interview Survey Data. Unpublished manuscript, American Statistical Association/National Center for Health Statistics, Washington, D.C.

Cohen, S. B. (1982) Comparison of design effect and relative variance curve strategy for variance estimation from complex survey data. Unpublished manuscript, National Center for Health Services Research. Presented at APHA, 1982.

Cox, B. G., & Cohen, S. B. (1985). Methodological Issues for Health Care Surveys. New York: Marcel Dekker, Inc.

Eargle, J., & U.S. Bureau of the Census (1990). Household wealth and asset ownership: 1988. Current Population Reports, Household Economic Studies, Series P-70(22), 1-47.

Hanson, R. F. (1978). The Current Population Survey: Design and methodology. Technical Paper (40). Washington, DC: United States Bureau of the Census.

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. Journal of the Royal Statistical Society, 36(1), 1-37.

Lee, E. S., Forthofer, R. N., & Lorimer, R. J.

(1990). Analyzing Complex Survey Data. Newbury Park, Calif.: Sage Publications.

Massey, J. T., Moore, T. F., Parsons, V. L., & Tadros, W. (1989). Design and Estimation for the National Health Survey, 1985-94. Vital Health Stat, 2(110), 1-33.

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

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

Table 1. Number of estimates of statistical quantities

Statistical quantity	Number of health coverage categories	Number of health status categories	Number of estimates: Possible	Number of estimates: Actual
1. Percent with insurance	12	107	1,284	1,038 *
2. Mean annual doctor visits	10	107	1,070	867 *
3. Mean hospital stay per discharge	10	107	1,070	724 *

* Excluding zero cells & repeats of categories

Table 2. Descriptive statistics for design effects (DEFF) for Direct Taylor Series and Relative Variance Curve Techniques

Statistical quantity	Direct Taylor series				Relative variance model			
	Mean	Med	skew-ness	CV	Mean	Med	skew-ness	CV
Percent with insurance	1.37	1.19	3.75	49.7	1.71	1.71	0.82	5.0
Mean annual physician contacts	1.14	1.11	1.73	24.4	0.79	0.55	2.14	97.8
Ratio hospital days to discharges	1.04	1.02	0.93	22.2	13.24	3.70	9.35	441.0

Table 3. Accuracy of alternative variance estimation techniques

Statistical quantity	Average Relative Absolute Difference and Standard Error		
	SRS	ADE	RELVAR
Percent with insurance	0.225	0.300	0.486
	(0.006)	(0.007)	(0.009)
Mean annual physician contacts	0.177	0.185	0.633
	(0.008)	(0.009)	(0.016)
Ratio hospital days to discharges	0.177	0.185	13.432
	(0.017)	(0.018)	(2.432)

ADE= average design effect model
RELVAR=relative variance curve model