EVALUATION OF DESIGN EFFECTS FOR THE MEXICAN AMERICAN PORTION OF THE HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY

Josefina A. Lago, Westat, Inc.
James Massey, Trena Ezzati, Clifford Johnson and Robinson Fulwood, National Center for Health Statistics
Josefina A. Lago, 1650 Research Blvd., Rockville, MD 20850

1. INTRODUCTION

The Hispanic Health and Nutrition Examination Survey (HHANES), conducted by the National Center for Health Statistics (NCHS) from 1982 to 1984, was the first large-scale survey to assess the health and nutritional status of the Hispanic population in the United States. Like earlier National Health and Nutrition Examination Surveys (NHANES), which focused on the general population, HHANES had an interview component and an examination component, that is, sample persons were interviewed in the household and examined in a mobile examination center (MEC).

The HHANES sample design was similar to the earlier NHANES, with two exceptions:

- Rather than a national sample, the universe for HHANES was restricted to the five Southwestern states in surveying the Mexican American population, to Dade County (Miami, FL) in surveying the Cuban-American population, and to New York City and nearby counties in surveying the Puerto Rican population.
- An average of three sample persons per household was selected in HHANES, while approximately one person per household was sampled in previous NHANES.

For each of the three populations of interest, the HHANES design was a multistage stratified cluster sample. The first stage units or primary sampling units (PSUs) consisted of counties with a high concentration or a high proportion of the particular Hispanic subgroup. The second stage units, area segments, consisted of a block or group of blocks defined within Census block groups or enumeration districts. To reduce the screening burden, block groups or enumeration districts with very few persons in the target population were considered out-of-scope. In sample segments, housing units were listed, subsampled, and screened; then eligible households were identified. A household was considered eligible for participation in the survey if at least one person identified him/herself as a member of the particular Hispanic subgroup. Depending on the household composition, all or a subsample of persons in eligible households were selected into the sample.

In computing design effects, as part of the preparation of public use tapes for the Southwestern component of HHANES, Kovar and Johnson[1] (1986) reported unexpectedly high (>7) and low (<.8) design effects for a number of statistics. This paper describes the results of research undertaken to investigate the possible reasons for such extreme variability in design effects. The design effects obtained with the linearization procedure SESUDAAN were first compared to those obtained by Balanced Repeated Replication (BRR) to see if instability of the variance estimates was a possible cause of the extreme variability. Since the results obtained by the two procedures were quite similar, the between- and within-PSU contributions to the total design effect were estimated and analyzed to identify possible sources of the large design effects. Also, the relationship between subdomain size and design effects was studied, and for statistics showing large between-PSU contributions to the design effect, the role of measurement error was investigated.

2. METHODOLOGY AND ANALYSIS

2.1 Comparison of Design Effects from BRR and SESUDAAN

The large design effects reported by Kovar and Johnson (1986) were mainly associated with the overall (across all sex and age classes defined for analyses purposes) and sex-specific estimates, but there were also large design effects for some sex-age subgroup estimates. One possible explanation for the large design effects is that SESUDAAN, the Taylor's series approximation procedure used by NCHS to compute HHANES variances, does not reflect the impact of the interview and examination nonresponse adjustments, noncoverage adjustment, and poststratification to sex-age population controls that were applied to the HHANES weights. To test this hypothesis, variances were computed on the same 21 data sets using the balanced repeated replication (BRR) procedure. The BRR procedure allows one to apply the same adjustments to the replicate weights that were applied to the overall weight.

For purposes of these analyses, we selected different types of health variables: body measurements, variables from the dental and physical examination, results from the biochemistry laboratory tests, and characteristics from the household interview. Some of the 21 variables used in these analyses are means and others are proportions. For each of the 21 data sets the definition of age-classes and analysis variables used in the BRR program was consistent with the specifications used by the NCHS analysts in computing the SESUDAAN variances.

The BRR procedure requires defining pseudo-strata[2] of two PSUs each. For the self-representing PSUs, since the area segments are the true first-stage sampling unit, pseudo-strata were formed by grouping the segments into two pseudo-PSUs. The segments were ordered in the sequence they were originally selected and the odd numbered segments were assigned to one pseudo-PSU while the even numbered segments were assigned to the other pseudo-PSU. For the nonself-representing PSUs, pseudo-strata were formed by pairing two similar PSUs (in terms of characteristics used in the PSU stratification). The PSU sample for the Southwest component of HHANES consisted of 14 PSUs, 12 nonself-representing PSUs and two self-representing, which resulted in 8 pseudo-strata.

As shown in Tables 1, 2, and 3, the design effects obtained by BRR and those obtained by the SESUDAAN procedure are generally quite close. However, for several variables such as height and weight where it is reasonable to expect that the post-stratification by age and sex had a significant impact, we found that the overall design effects obtained with BRR are much smaller than those obtained by SESUDAAN. To summarizes these results, column (a) in Table 4 shows the correlations between the BRR design effects and the SESUDAAN design effects. These correlations certainly support the conclusion that the two variance estimation procedures yield similar results whenever the effect of poststratification is negligible.

2.2 Calculation of the Between-PSU and Within-PSU Components of the Total Design Effect

After determining that the extreme variability in design effects were neither due to instability in the variances nor to the fact that the adjustments applied to the HHANES weight were not reflected by SESUDAAN, we looked into components of variance and the contributions to the total design effect from the between-PSU and within-PSU variability.

For an estimated mean or proportion, the variance of the estimate may be expressed as

$$\sigma^2(y') = \sigma_B^2(y') + \sigma_W^2(y')$$

where

$\sigma^2(y')$ = the total variance of the estimate;

$\sigma_B^2(y')$ = the between-PSU component of variance;

$\sigma_W^2(y')$ = the within-PSU component of variance.

The between-PSU component of variance reflects the contribution to variance due to sampling PSUs. The within-PSU component reflects the variability arising from the selection of segments within PSUs, households within segments, and persons within households. The within-PSU component also reflects the additional variability resulting from the differential sampling rates used to sample certain age groups. It should be noted, however, that for HHANES analyses carried out by age-sex subdomains, the effect of clustering by household will be trivial since a sample household will usually provide at most one person in a given subdomain. Similarly, the overall sampling rate will be the same for all persons in sex-age subdomain.

Estimates of the components of variance were computed using BRR. Depending on how the pairs (half-samples) within pseudo-strata are defined, BRR can be used to estimate components of variance. To estimate the total variance, $\sigma^2(y')$, the assignment of units within each pseudo-stratum was made by pairing PSUs in the case of noncertainty pseudo-strata, and pairing segments in the case of certainty strata. To estimate the within-PSU variances, the pairing was done by segments in all strata.

After computing the total and within-PSU variance component as described above, we derived the total design effect $(\sigma^2(y')/\sigma^2_{SRS})$ and within-PSU design effect $(\sigma^2_{w}(y')/\sigma^2_{SRS})$. The between-PSU and within-PSU contributions to the total design effect were then estimated as described below.

Let

D = total design effect

$\bar{m}$ = he average number of sample segments per sample PSU;

$\bar{\bar{n}}$ = the average of sample households per sample segment;

$\bar{\bar{\bar{q}}}$ = the average number of sample persons per sample household;

ρ_1 = the intraclass correlation between examined persons within PSUs;

ρ_2 = the intraclass correlation between examined persons within segments.

An expression for D which is useful for examining the contributions to the total design effect of the different sources of variability is:

$$\begin{aligned} D &= 1 + \rho_1(\bar{m}\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1) + \rho_2(\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1) \qquad (1) \\ &= 1 + D_1 + D_2 \end{aligned}$$

that is

$$\begin{aligned} D &= \rho_1(\bar{m}\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1) \\ &= \text{between-PSU contribution to the design effect} \end{aligned}$$

and

$$\begin{aligned} D_2 &= \rho_2(\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1) \\ &= \text{within-PSU contribution to the design effect.} \end{aligned}$$

In Table 5 the columns headed "between contribution (D_1)" and "within contribution (D_2)" show the contributions to the total design effect due to sampling PSUs and within-PSU sampling respectively. The results presented in Table 5 show that for estimates with large or moderately large design effects, the dominating contribution to the design effect is associated with the sampling of PSUs (D_1). (Of course, for estimates that have design effects less than one or close to one this conclusion does not apply.) We recognize that the small number (14) of PSUs in HHANES may be a key contributing factor to the large between-PSU contribution to the total design effect. However, for a survey like HHANES, where three target populations were sampled in different parts of the country and which required moving the MECs between PSUs to carry out the examination component of the study, it was not possible to have a much larger PSU sample from the cost stand point.

To better understand the results presented in Table 5, it should be noted that often the estimated within-PSU design effect is less than one. This may indicate that the implemented within-PSU sample design is more efficient than simple random sampling but also could be a result of instability in the variance estimate. In these cases we have set D_2 equal to zero before computing the averages shown in Table 5. Another result that may be due to instability in the variance estimates is when the within-PSU variance computed by BRR is greater than the BRR estimate of total variance, which results in a negative between-PSU variance and thus a negative within PSU contribution to the design effect. In this case we set to zero the between-PSU contribution to design effect, D_1 before computing the averages shown in Table 5.

2.3 Relationship between Design Effects and Subdomain Sizes

In reviewing the results shown in Tables 1 to 3, it is clear that the design effects associated with overall, sex-only estimates are generally larger than those by sex-age subdomain. In trying to understand the possible reasons for these results, it is useful to refer to expression (1) in Section 2.2 for the total design effect (D), namely

$$D = 1 + \rho_1(\bar{m}\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1) + \rho_2(\bar{\bar{n}}\bar{\bar{\bar{q}}} - 1)$$

It is easy to see that if the average number of sample persons per PSU $(\bar{m}\bar{\bar{n}}\bar{\bar{\bar{q}}})$ or per segment $(\bar{\bar{n}}\bar{\bar{\bar{q}}})$ for the overall and sex-only estimates is about 10 times larger than the corresponding number of sample person in each individual sex-age subdomain, even relatively small interclass correlations, ρ_1 and, ρ_2 would result in relatively larger design effects for the overall estimates than for the sex-age subdomain estimates. It should be noted that for characteristics that are significantly affected by the poststratification imposed on the final HHANES weights, taking this into account in the variance estimation procedure results in reasonable design effects even for the larger subdomains (see Section 2.1).

To corroborate the conclusion that design effects are correlated with sample size, Table 4 (columns (b) and (c)) shows the correlation coefficient ($r_{x,y}$) between sample size and design effects (both BRR and SESUDAAN) for the 21 variables under consideration. For each variable the correlation was obtained by treating each subdomain's sample size and total design effect as the pair of values (x_i,y_i). Except for a few variables -- those that have rather small design effects for all estimates -- the correlation is quite high. For half of the variables studied, the correlation coefficient between the BRR design effect and the sample size is greater than .8. The results for SESUDAAN are very similar.

Since most of the analysis performed on the HHANES data are by sex-age subdomain, these average design effects are of greatest interest. As shown on Tables 2 and 3 in the maximum and minimum columns, there are subdomain design effects as high as 10.6 and as low as .08. The averages, however, are well behaved and range between .97 and 3.66. This result suggests that in estimating variances for HHANES, estimates using average design effects may be preferable to point estimates of variance.

2.4 The Role of Measurement Variability

To investigate the possible role of measurement error associated with the two MEC examination teams used in HHANES, we ran t-tests comparing the MEC team means, for a few of the examination variables that had large between-PSU contribution to the design effect and showed considerable variability among unweighted PSU means. In these t-tests we treated the PSU unweighted means as the analysis unit and the examination team as the grouping factor. The t-statistic was used to detect significant differences between the teams. Although we realize that treating this problem as a t-test problem ignores some of the theoretical assumptions of the statistic, we felt it was a useful way of looking at the role of measurement variability.

Table 6 shows the results to the t-tests for the selected variables. Although the results for these few variables are far from conclusive, they indicate that the large between-PSU contribution to the total variance cannot be attributed to differences between the examination teams (with the exception of mean debri index). They suggest that true between-PSU variability dominates the variance .

3. SUMMARY AND RECOMMENDATIONS

The results of our research can be summarized as follows:

- The estimates of design effects obtained with BRR are quite similar to those obtained by SESUDAAN, except for variables such as height and weight where it is reasonable to expect that the poststratification by age and sex had a significant impact.
- The magnitude of the design effect is correlated with sample size; that is, design effects tend to be larger for overall and sex-only domains than for sex-age subdomains.
- Average design effects for the individual age-sex subdomains, across groups of variables, tend to be well behaved.
- For estimates with large total design effects, the between-PSU contribution to the total design effect seems to dominate. Moreover, the large between-PSU contribution to the total design effect does not appear to be due to MEC team variability.

The research we conducted suggests several areas in the design of large multistage surveys that should be studied further.

- Since average design effects, across groups of variables, for individual sex-age subdomains, are fairly well behaved this suggests the use of generalized variances in place of point estimates of variance
- Although a larger PSU sample may reduce the large contribution of the sampling of PSUs to the total design effect, this is not a realistic option for surveys like HHANES which focus on a minority population concentrated on a particular area of the country. The role of between-team variability in the between PSU variance should be investigated further.
- Some studies comparing the jackknife procedure for variance computation to the BRR and Taylor's linearization suggest that for estimates from complex surveys with a small number of PSUs, the jackknife variance estimation procedure result in lower estimates of variance. This result could be explored with the three HHANES datasets.

FOOTNOTES

[1] M. G. Kovar and C. Johnson, Design Effects from the Mexican American Portion of the Hispanic Health and Nutrition Examination Survey, presented at the Survey Research Section, ASA, August 1986.

[2] We refer to pseudo-strata because the HHANES design is a 1-PSU-per-stratum design. Strata were collapsed to form pseudo-strata.

Table 1. Comparison of BRR and SESUDAAN average, design effects for sex-age subdomain, sex-only subdomains and overall

STATISTIC	Average DEFFT for age-sex subdomains		Average DEFFT for sex-only subdomains		DEFFT for overall estimate	
	BRR	*SESDN*	*BRR*	*SESDN*	*BRR*	*SESDN*
BODY MEASURES						
Mean Subscapular Skinfold	1.94	1.82	5.99	4.92	9.75	7.53
Mean Weight	1.24	1.16	0.52	1.66	0.64	2.31
Mean Height	1.58	1.57	0.80	2.62	1.20	3.59
DENTAL EXAM						
Proportion with Upper Denture	1.57	1.51	1.53	2.04	2.25	3.13
Mean Debris Index	3.65	3.53	13.56	12.77	26.00	24.50
Prop. w. Previous Ortho Treatment	1.46	1.33	5.28	4.38	9.35	8.08
Mean Calculus Index	1.70	1.58	4.03	3.07	7.39	4.97
PHYSICIAN EXAM						
Mean Systolic Blood Pressure	2.13	2.02	4.86	4.46	8.36	7.93
Proportion with Heart Murmur	1.92	1.80	4.65	5.21	8.52	9.67
Proportion with Strabismus	3.66	3.28	18.73	16.60	35.69	31.50
Proportion With Scoliosis	3.54	3.43	22.1	22.20	43.40	43.60
ADULT SP QUESTIONNAIRE						
Prop. Covered by Dental Insurance	1.64	1.58	3.81	3.57	7.38	6.86
Proportion Ever Done Farm Work	2.33	2.21	8.30	7.62	14.25	12.40
Proportion w. Glasses or Contacts	1.38	1.36	2.32	2.71	3.93	4.66
Mean Age First Smoked Cigarettes	1.04	0.94	0.87	0.99	0.34	0.48
Proportion Usual Place Health Care	1.16	1.05	1.78	1.88	3.29	3.34
BIOCHEMISTRY						
Mean Serum Cholesterol	0.97	0.96	0.80	1.09	0.65	1.49
Mean Lead	1.37	1.59	3.07	2.45	2.60	2.83
Mean Hemoglobin	3.57	3.42	20.10	19.18	25.40	19.88
Mean Corpuscular Hemoglobin	1.62	1.62	4.05	4.78	7.06	7.39
Mean Serum Iron	1.22	1.14	2.07	1.93	2.94	2.55

Table 2. Average, maximum, and minimum design effects for hematology /biochemistry variables, overall and by age-sex class

Sex	Age	BRR Average DEFFT	BRR Maximum DEFFT	BRR Minimum DEFFT
All	All	7.73	25.40	0.34
Male	All	3.41	9.19	0.96
Female	All	8.63	31.02	0.63
Male	4-5	1.46	2.89	0.42
	6-8	1.63	3.23	0.50
	9-11	1.99	4.06	0.66
	12-17	2.02	4.16	0.80
	18-24	1.85	3.23	0.69
	25-34	1.81	4.72	0.51
	35-44	1.22	1.60	0.50
	45-54	1.21	1.68	0.70
	55-64	1.46	2.76	0.89
	65-74	1.06	1.82	0.15
Female	4-5	2.19	2.90	1.48
	6-8	1.41	3.06	0.41
	9-11	3.18	5.66	1.59
	12-17	2.88	6.42	0.60
	18-24	2.67	5.62	0.99
	25-34	1.90	4.93	0.46
	35-44	2.60	5.20	1.10
	45-54	1.95	3.77	0.98
	55-64	1.92	3.08	0.54
	65-74	1.62	4.42	0.24
All	4-5	1.86	3.84	0.66
	6-8	2.17	5.31	0.47
	9-11	4.36	8.78	1.16
	12-17	3.11	7.12	0.88
	18-24	2.05	3.39	0.60
	25-34	1.75	4.62	0.33
	35-44	1.56	2.35	0.74
	45-54	1.87	2.81	0.80
	55-64	1.56	2.11	1.04
	65-74	1.58	3.76	0.73

Table 3. Average, maximum, and minimum design effect for all other variables, overall and by age-sex class

Sex	Age	BRR Average DEFFT	BRR Maximum DEFFT	BRR Minimum DEFFT
All	All	11.37	43.40	0.34
Male	All	5.88	21.91	0.47
Female	All	6.50	22.18	0.56
Male	6mth - 5 yr	1.56	3.43	0.61
	6-11	2.47	4.99	0.65
	12-17	2.86	6.67	0.85
	18-24	1.82	5.05	0.27
	25-34	2.11	5.86	0.38
	35-44	1.27	2.46	0.12
	45-54	1.97	8.05	0.08
	55-64	1.70	3.09	0.30
	65-74	1.39	3.45	0.32
Female	6mth - 5 yr	2.03	7.87	0.50
	6-11	2.22	6.22	0.78
	12-17	1.89	4.33	0.60
	18-24	2.94	5.92	0.61
	25-34	2.83	9.05	0.25
	35-44	1.88	6.72	0.47
	45-54	2.12	5.99	0.86
	55-64	1.83	3.67	0.55
	65-74	1.38	2.15	0.64
All	6mth - 5 yr	2.45	7.83	0.57
	6-11	3.73	9.92	0.94
	12-17	3.72	8.46	1.38
	18-24	3.57	10.65	0.93
	25-34	3.67	10.85	0.40
	35-44	2.14	6.59	0.35
	45-54	2.97	11.38	0.42
	55-64	1.98	3.88	0.45
	65-74	1.73	3.61	0.49

Table 4. Correlation between BRR and SESUDAAN design effects and between sample size and design effects

STATISTIC	Correlation Coefficients		
	BRR Deft vs. SESUDAAN	Sample Size vs. BRR Deft.	Sample Size vs. SESUDAAN
BODY MEASURES			
Mean Subscapular Skinfold	0.99	0.87	0.83
Mean Weight	0.67	-0.29	0.45
Mean Height	0.49	-0.22	0.58
DENTAL EXAM			
Proportion with Upper Denture	0.85	0.28	0.49
Mean Debris Index	0.86	0.90	0.77
Prop. w. Previous Ortho Treatment	0.89	0.88	0.73
Mean Calculus Index	0.95	0.81	0.70
PHYSICIAN EXAM			
Mean Systolic Blood Pressure	0.99	0.84	0.82
Proportion with Heart Murmur	0.99	0.79	0.83
Proportion with Strabismus	0.99	0.94	0.95
Proportion With Scoliosis	0.99	0.96	0.96
ADULT SP QUESTIONNAIRE			
Prop. Covered by Dental Insurance	0.99	0.83	0.81
Proportion Ever Done Farm Work	0.99	0.94	0.95
Proportion w. Glasses or Contacts	0.98	0.40	0.53
Mean Age First Smoked Cigarettes	0.90	-0.35	-0.29
Proportion Usual Place Health Care	0.97	0.60	0.65
BIOCHEMISTRY			
Mean Serum Cholesterol	0.87	-0.20	0.26
Mean Lead	0.85	-0.03	0.34
Mean Hemoglobin	0.99	0.86	0.84
Mean Corpuscular Hemoglobin	0.97	0.81	0.82
Mean Serum Iron	0.98	0.46	0.47

Table 5. Estimated average contributions to the design effect for sampling PSUs and sampling within-PSU, overall, by sex, and sex and age class

STATISTIC	Overall		Sex-only average		Sex-age average	
	Between contribution (D1)	Within contribution (D2)	Between contribution (D1)	Within contribution (D2)	Between contribution (D1)	Within contribution (D2)
BODY MEASURES						
Mean Subscapular Skinfold	8.73	0.02	4.99	0.00	0.78	0.28
Mean Weight	0.00	0.00	0.00	0.00	1.16	0.25
Mean Height	0.20	0.00	0.01	0.00	0.56	0.23
DENTAL EXAM						
Proportion with Upper Denture	1.25	0.00	0.56	0.00	0.16	0.24
Mean Debris Index	23.55	1.45	11.70	1.01	2.24	0.33
Prop. w. Previous Ortho Treatment	6.32	2.03	3.11	1.17	0.48	0.16
Mean Calculus Index	6.05	0.34	3.03	0.00	0.58	0.16
PHYSICIAN EXAM						
Mean Systolic Blood Pressure	7.56	0.00	3.86	0.00	1.07	0.09
Proportion with Heart Murmur	6.93	0.59	3.30	0.35	0.76	0.32
Proportion with Strabismus	33.49	1.20	17.07	0.66	2.51	0.21
Proportion With Scoliosis	41.59	0.81	20.77	0.25	2.00	0.58
ADULT SP QUESTIONNAIRE						
Prop. Covered by Dental Insurance	5.60	0.78	2.67	0.15	0.43	0.33
Proportion Ever Done Farm Work	12.46	0.79	6.46	0.84	0.94	0.46
Proportion w. Glasses or Contacts	2.77	0.16	1.01	0.32	0.33	0.34
Mean Age First Smoked Cigarettes	0.00	0.00	0.00	0.00	0.17	0.18
Proportion Usual Place Health Care	1.24	1.05	0.39	0.40	0.07	0.34
BIOCHEMISTRY						
Mean Cholesterol	0.00	0.00	0.00	0.00	0.11	0.13
Mean Lead Value	1.36	0.26	1.99	0.08	0.53	0.27
Mean Hemoglobin	24.40	0.00	18.87	0.24	2.34	0.23
Mean Corpuscular Hemoglobin	4.03	2.03	2.26	0.76	0.39	0.37
Mean Serum Iron	1.94	0.00	1.05	0.02	0.25	0.20

Table 6. Results of ANOVA for selected variables with high design effects and showing considerable variability among PSU means

Statistic	Overall design effect	Mean for MEC team 1	Mean for MEC team 2	t-test	Significance level
Proportion w. scoliosis	43.3	.011	.013	1.75	.103
Proportion w. strabismus	35.7	.046	.133	1.65	.121
Mean subscapular skinfold	9.8	16.07	15.98	.116	.909
Proportion w. heart murmur	8.5	.046	.041	.217	.831
Mean debri index	26.0	.630	.970	6.17	.0001 *
Mean calculus index	7.4	43.8	46.8	.452	.658

(*) The MEC teams are significantly different