DESIGN EFFECTS FROM THE MEXICAN AMERICAN PORTION OF THE HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY: A STRATEGY FOR ANALYSTS

Mary Grace Kovar and Clifford Johnson
National Center for Health Statistics

Introduction

In the data from the Hispanic Health and Nutrition Examination Survey, we found design effects that were less than one and others that were greater than seven. Different strategies for stratification or incorporating the third level of sampling did not solve the problem, nor did using a different approach to estimating variances. The design effects were not associated with the size of the sample in the cells.

We developed a strategy of using average design effects to cope with the problem so that the public-use data tapes could be released and analysis could proceed. We do not regard this strategy as a definitive solution.

Moreover, further investigation revealed that the problem was not confined to the Hispanic HANES, although there was a larger proportion of both large and small design effects in that survey than in the earlier National HANES. We think that the problem may be more widespread than is generally realized and that further work is needed.

Background

The Hispanic Health and Nutrition Examination Survey (Hispanic HANES or HHANES) is the most recent of a series of Health and Nutrition Examination Surveys conducted by the National Center for Health Statistics. These surveys are unique in that the persons selected in the sample are both interviewed in the household and examined in mobile examination centers that are moved from one sampling point to another. The logistics of the survey mean that the number of sampling points must be limited and a sufficient number of people examined at each site to achieve the desired number in the total sample. Therefore, the surveys are characterized by a relatively small number of PSU's with a relatively large number of persons in each PSU (1,2).

The Hispanic HANES sample design was similar to those of the previous National Health and Nutrition Examination Surveys (National HANES or NHANES). All of the surveys have used complex, multistage, stratified, clustered samples of defined populations. In hierarchal order, the stages of selection for each survey were primary sampling units (PSU's), segments consisting of clusters of households, households, and eligible persons in the households.

One difference is that, unlike the National HANES, the sample was designed to have more than one sample person from each family. Since most analyses of HANES data are age-specific, this does not increase family clustering. It does, however, permit family analyses that were not possible in previous surveys.

A second difference is that, while the National HANES sample had 8 households and 8 sample persons per segment, the Hispanic HANES had 6 households and 18 sample persons per segment. Thus, in the National HANES there was an average of 54 segments in each primary sampling unit and in the Hispanic HANES there were 39.

The third difference between the Hispanic HANES and the National HANES was that the Hispanic HANES was a survey of three special population subgroups in selected areas rather than a national sample. The three subgroups are: Mexican Americans in five southwestern states (Arizona, California, Colorado, New Mexico, and Texas), Cuban Americans (Dade Co., Florida, and Puerto Ricans (New York metropolian area). The data in this paper are from the Mexican American sample only.

The HHANES Mexican American sampling frame was restricted to counties that had sufficient Mexican Americans to make the survey economically possible. The frame consisted of 193 PSU's that contained 84 percent of the Mexican American population of the United States as enumerated in the 1980 Census and 97 percent of those in the five states (3).

Information from the 1980 Census was used to stratify the PSU's into 14 strata of approximately equal size. Two counties - Los Angeles CA and Bexar TX - were certainty strata and were selected with a probability of one. One PSU was selected from each of the 12 non-certainty strata.

Because Mexican Americans were a minority in most PSU's and screening costs were high, block groups and enumeration districts with very few Hispanics enumerated on the 1980 Census were considered out-of-scope as were institutions, Indian reservations, and military bases. After those eliminations, secondary sampling units were selected with the sizes chosen to produce about 18 "eligible" Hispanics (age 6 months -74 years and self-identifying as Mexican American) in each secondary unit.

An "eligible" household was one with at least one Hispanic person ages 6 months - 74 years. In households containing at least one "eligible" Hispanic, everyone else who was 6 months - 74 years of age was also eligible for the survey. Approximately 3/4 of those ages 6 months - 19 years, 1/2 of those ages 20-44 years, and all those age 45 years and older were selected.

In order to make population estimates and calculate variances that incorporated the complex sample design, weights were calculated for each survey component for each person and a pseudo design of 2 PSU's per stratum was created (4).

The weights for each person characteristic were the product of:

1. Reciprocals of the probability of selection at each stage of the design,
2. Adjustments for interview and examination nonresponse within homogeneous socio-demographic cells,
3. Adjustment for noncoverage within the sample PSU's, and
4. Poststratified ratio adjustment by age and sex to current Bureau of the Census estimates of the civilian non-institutionalized target population.

The 2 PSU's per stratum design was created by:

1. Splitting each of the 2 certainty strata into 2 pseudo-PSU's, and
2. Pairing the 12 PSU's from the non-certainty strata into 6 pseudo-strata using the stratification criteria topair "like" PSU's.

The number of selected, interviewed, and examined persons is shown for each of the strata and PSU's in this design in Table 1.

The Problem

As part of the editing of the Hispanic HANES data for release on public-use data tapes, point estimates and their corresponding variances were calculated for many of the survey components using the SESUDAAN procedure (5).

One of our first observations was that there were a number of unexpectedly large (>7.5) and small (<0.8) design effects, where the design effect is defined as the ratio of the variance from the complex sample to the variance from a simple random sample of the same size. The distribution varied for each data component that was examined. Some had many small design effects, others had many large ones, and still others had both. There seemed to be little pattern. For one component, the design effect might be large for one age group while for an adjoining one - with approximately the same number of sample persons in the cell - it might be less than one. For a different component, the distribution of design effects by age groups was dissimilar.

Because of previous concerns (6) about the construction of the PSU pairings, we investigated alternative strategies for creating the PSU's and strata. The alternative approach was to re-stratify into 4 strata each with 4 PSU's using the same criteria for pairing that had been used originally. The estimated variances were different, but they were just as erratic.

A second alternative approach was coding the secondary sampling units and including them as the third level of selection in SESUDAAN. This approach was equally non-productive. There was no major change in the variance estimates or in the distribution of the design effects. For some of the survey components, these approaches resulted in an even wider range of design effects.

We also reasoned that the findings might be the result of some problem specific to SESUDAAN. Therefore, the estimated variances were recalculated using the balanced repeated replication program developed at NCHS for use in the health examination surveys (7). This program, unlike SESUDAAN, takes poststratification into account. The variances were virtually identical with those from SESUDAAN.

The variances from the HHANES were erratic. The problem did not appear to be a result of the method of estimating variances. It was not alleviated by restratification or by incorporating the third level of sampling. They were so erratic that we were concerned that analysts who followed our recommendations to estimate variances for the individual estimates in all inferences or hypothesis testing (8,9) would draw incorrect conclusions. We did not believe that the design effects for adjacent age groups were, say, 3.0, 0.3. 0.7, and 7.0 but the analyst who did not take care would be likely to conclude that there was a difference between the second and third but not between the first and second or third and fourth.

Therefore, we had to develop some recommendations for the analysts - especially those who would be using the public-use data tapes. The recommendation was to consider using average design effects to correct variances based on the assumption of simple random sampling.

Examples

Since the advantages in programming and computer time were considerable if we used SESUDAAN (8), and there were no apparent advantages in using either of the alternative sample design approaches or in using BRR rather than a Taylor series approach to estimating variances, we have used SESUDAAN and the original design of 8 strata each containing 2 PSU's.

The three measures we are focusing on are the body measures of height and weight and the blood measurement of hemoglobin. By using these, we hoped:
1. to eliminate the possibility of confounding by interviewer variation because the information was collected in the mobile examination centers under standardized conditions,
2. to choose variables for which everyone had a positive, i.e., a non-zero response, and
3. to make comparisons with another survey.

Data on these three measures were collected in the NHANES II using exactly the same procedures. The NHANES II, conducted during 1976-80 was a national sample with a design similar to the HHANES design. However, there were 64 pseudo-PSU's in 32 pseudo-strata. That gave us the opportunity to compare the distribution of the design effects for the same variables collected in the same manner on two surveys that differed, for this purpose, only in the number of sampling points.

The design effects were calculated for each of the age-sex groups used in the Vital and Health Statistics Series 11 publications and then averaged. As you can see from Table 2, the average design effects were generally, but not always, larger for the HHANES than for the NHANES II. The average design effects were larger for hemoglobin than for either of the body measures in both HHANES and NHANES II.

The size of the average design effect tells you little about the distribution, however, and as you can see from Table 3 the distributions were generally more concentrated in the NHANES II that had 32 PSU's than in HHANES with 16. The number of large design effects is particularly noticeable for hemoglobin; some of the design effects were below one for each of the body measures.

Overall, for the 15 biochemical and hematological assessments for which we have data from both surveys, 5.3 percent of the design effects from NHANES II and 18.0 percent of those from HHANES were below one; 4.8 percent of those from NHANES II and 13.6 percent of those from HHANES were 5 or higher. The problem of unexpectedly large or small design effects is not confined to the smaller HHANES, but it is more prevalent.

We anticipated that the design effects might be a function of the number of sample persons in a cell -that the unusually large or small design effects might be concentrated in cells with few sample persons. As you can see from Table 4, they weren't. The range of design effects in cells with 500 or more sample persons was from less than 0.5 to greater than 5.0. The smallest cells did not have unusually large design effects.

The ranges for other variables show the same kinds of distributions. For some of the interview data the design effects appear to be randomly distributed between 0.5 and about 7.5. There are extremely large design effects for some of the interview variables and small ones for others. In general, there is no pattern that we can discern.

Discussion

Even though the total number of examined persons in the Mexican American portion of the Hispanic HANES is large, there are few instances in which the total sample is the group of interest. The analytic interest is almost always in subclasses. Subclass analysis can lead to estimates that are unstable, particularly estimates of variances. Consequently, analyses of subclasses require that users pay particular attention to the number of sample persons in the subclass and the number of PSU's that contain at least one sample person in the subclass. Small sample sizes, or a small number of PSU's used in the variance estimation, may produce unstable estimates of the variances.

The analytic strategy that we have been

recommending for the Health and Nutrition Examination Surveys has been to do preliminary examinations of the data without the weights and assuming simple random sampling but to base all final analyses on the weighted data incorporating the variance estimates based on the complex sample design. In addition, we have recommended incorporating the variance-covariance matrix that incorporated the complex sample design in all hypothesis tests (9,10).

However, having found that the variances estimated from the Mexican American portion of the Hispanic HANES are so unstable, we have begun looking at the variances from the earlier NHANES more carefully, and have found that they were also more unstable than we had realized. We think that the strategy must be re-examined.

Our long-term strategy is to develop better analytic software that incorporates the multi-stage sample and the post-stratification.

Our short-term strategy is to present illustrative average design effects as part of the public-use data tape documentation. The averages are based on the age groups used in the Vital and Health Statistics Series 11 publications and are given for both sexes and for each sex separately. The design effects in the documentation give the user an idea of the range - even in the averages - for selected response variables on each tape. There is also an example of how to use the average design effects.

It is not a very satisfactory solution. It adds to the already large burden of preparing the public-use tapes from a survey that may result in 10-15 such tapes. It cannot include design effects for all of the many subclasses that analysts may choose to use in their own work; the most we can do is include the age-sex groups that are standard in almost all work. It does not solve the problem for analysts who wish to incorporate the co-variances unless they choose to multiply the entire matrix by the average design effect. As part of the research leading to NHANES III, which is scheduled to begin in 1988, we are seeking better solutions. We welcome your criticisms, comments, and suggestions.

Acknowledgments

Cecilia Snowden and Gretchen Jones, Office of Research Methods, NCHS, did the programming to investigate whether alternative approaches to estimating variances would alleviate the problem. Robinson Fulwood and Michael Rowland, Division of Health Examination Statistics NCHS, provided much of the data used in this paper. Many analysts in DHES used each of the 4 approaches to examine a specific component of the data. Robert Casady provided invaluable support in suggesting approaches and solutions.

References

1. National Center for Health Statistics: Plan and Operation of the Second National Health and Nutrition Examination Survey, 1976-80. Vital and Health Statistics. Series 1 No 15. DHHS Pub. No. (PHS) 81-1317. Public Health Service. Washington. U.S. Government Printing Office.

2. National Center for Health Statistics: Maurer, KR and others: Plan and Operation of the Hispanic Health and Nutrition Examination Survey, 1982-84. Vital and Health Statistics. Series 1 No 19. DHHS Pub. No. (PHS) 85-1321. Public Health Service. Washington. U.S. Government Printing Office.

3. Gonzales, JF, AA White, and T Ezzati: Sample Design for the Hispanic Health and Nutrition Examination Survey, 1982-84. Proceedings of the Survey Research Methods Section, American Statistical Association, 1984.

4. Gonzales, JF and T Ezzati: Estimation in the Southwest Component of the Hispanic Health and Nutrition Examination Survey, Proceedings of the Survey Research Methods Section, American Statistical Association, 1985.

5. Shah, BV: SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. RTI/5250/00-01S. Research Triangle Institute, Research Triangle Park, NC, 1981.

6. Gonzales memorandum.

7. Jones, G. The NCHS BRR program.

8. Cohen, S, V Burt, and G Jones: Efficiencies in Variance Estimation for Complex Survey Data. American Statistician, May 1986.

9. Kovar, MG: Chapter 11 in reference 2.

10. Landis, JR, JM Lepkowski, SA Eklund, and SA Stehouwer: A Statistical Methodology for Analyzing Data from a Complex Survey: The first National Health and Nutrition Examination Survey. Vital and Health Statistics. Series 2, No 92. DHHS Pub. No. (PHS) 82-1366. Public Health Service. Washington.U.S. Government Printing Office.

Table 1. Number of selected, interviewed, and examined persons in each PSU. Hispanic HANES, Mexican American sample

		Number of persons		
Stratum	PSU	Sample	Interviewed	Examined
1	1 2	(2,449)	(2,106)	(1,812)
2	3 4	(505)	(404)	(341)
3	5	626	542	481
	6	649	556	490
4	7	494	419	336
	8	439	356	299
5	9	607	533	456
	10	654	596	523
6	11	555	517	480
	12	521	459	417
7	13	584	513	491
	14	614	555	494
8	15	553	425	356
	16	644	573	486
Total		9,894	8,554	7,462

Table 2. Average design effects for selected variables measured on examination.
Hispanic HANES, Mexican American sample and NHANES II

		Total	Males	Females
Hemoglobin	HHANES	4.1	2.8	4.2
	NHANES II	2.8	2.2	2.2
Height	HHANES	1.7	1.5	1.9
	NHANES II	1.4	1.7	1.4
Weight	HHANES	1.5	1.3	1.1
	NHANES II	1.4	1.4	1.4

Averages of analytic age groups.

Table 3. Distribution of design effects for hemoglobin, height, and weight for analytic age groups.
Hispanic HANES, Mexican American sample and NHANES II sample

Design Effect	Hemoglobin HHANES	Hemoglobin NHANES II	Height HHANES	Height NHANES II	Weight HHANES	Weight NHANES II
0.5			2			
0.5-0.99			4	4	9	3
1.0-1.49	3	4	5	11	9	15
1.5-1.99	2	7	9	8	6	7
2.0-2.49	4	8	3	3	2	1
2.5-2.99	3	6	2	1	1	1
3.0-3.49	1	2	1			
3.5-3.99	3		1			
4.0-4.49	5					
4.5-4.99	2	1				
5.0 +	7					

Table 4. Distribution of design effects for hemoglobin, height, and weight by sample size for analytic age groups.
Hispanic HANES, Mexican American sample

Design Effect	Sample size 50-99	100-149	150-199	200-249	250-299	300-349	350-399	400-449	450-499	500 plus
<0.5	1									1
0.5-0.99	1				2		2	2		6
1.0-1.49		1		3	3		1	1	1	7
1.5-1.99	1	1		4		2		1		8
2.0-2.49				2	1	1		2	1	2
2.5-2.99					1	1			1	3
3.0-3.49									1	1
3.5-3.99		1			1		1			1
4.0-4.49					1		1		1	2
4.5-4.99								1		1
5.0 +						1			1	5