NONRESPONSE AND NONCOVERAGE ANALYSIS IN THE SOUTHWEST COMPONENT OF THE HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY

Joe Fred Gonzalez, Jr. and Marilyn McMillen
National Center for Health Statistics

Introduction

The Hispanic Health and Nutrition Examination Survey (HHANES), sponsored by the National Center for Health Statistics (NCHS), was the first large scale multistage probability sample survey to assess the health and nutritional status of Hispanics in the United States. The HHANES was a multi-purpose survey consisting of an examination by a physician, a dental examination, various physiological measurements and laboratory tests. The HHANES was a subnational survey and consisted of three separate target populations: persons 6 months to 74 years of age and of Mexican origin residing in the Southwest (Arizona, California, Colorado, New Mexico, and Texas); persons in the same age group who were of Cuban origin residing in Dade County (Miami, Florida); and persons in the same age group who were of Puerto Rican origin residing in the New York City area. Separate estimates will be produced for each of the three target populations. This paper will focus on nonresponse and noncoverage as possible sources of bias in the Southwest HHANES estimates.

Sample Design of the Southwest HHANES

The following is a brief description of the sample design of the Southwest HHANES. A more detailed description of the HHANES sample design can be found in an NCHS publication [1] and in two previous papers [2,3]. Although the general structure of the HHANES sample design and operation was similar to both of NCHS' first National Health and Nutrition Examination Survey (NHANES I) [4] and the second National Health and Nutrition Examination Surveys (NHANES II) [5], there was a major difference between the HHANES and these previous NCHS surveys. The HHANES was a subnational survey of a special subgroup of the U.S. population.

For the Southwest HHANES, a complex, multistage, stratified, probability cluster design was used to survey persons of Mexican origin. The four stages of selection were primary sampling units or PSUs (counties or small groups of contiguous counties), segments (clusters of households), households, and persons. The sampling units at the PSU and segment stage were stratified prior to selection. The sample was designed to be approximately (though not exactly) self-weighting, i.e., efforts were made in the selection process to assure that every Hispanic sample person had approximately the same probability of selection.

Although the target population for the Southwest HHANES was conceived to be all households with at least one member of Mexican origin, sampling and data collection were restricted to counties that had a sufficient number and/or percentage of Hispanics to render it economically feasible to establish and operate a medical examination center (MEC) over a 4 to 7 week time period [6]. For purposes of sampling and data collection the Southwest HHANES universe consisted of 193 PSUs which included about 84

percent of the 1980 Mexican origin population in the United States and about 97 percent of the 1980 Mexican origin population in the five southwestern states. The net coverage rate of the 1980 Mexican origin population in the Southwest was approximately 90 percent. As will be seen later, one of the goals of the estimation procedure was to adjust the observed data to compensate as much as possible for the undercovered population.

Estimation Methods

Estimates for the Southwest HHANES were derived through a multistage estimation procedure which was designed to yield statistics that come close to minimizing the mean square errors of desired estimates. A detailed description of the Southwest estimator is given in [7]. The procedure had four basic features and the final weight associated with an examined sample person was the product of the following four components:

1. inflation of sample person observations by the reciprocals of the probabilities of selection at each stage of the design: PSU, segment, household, and sample person;

2. adjustments for interview and examination nonresponse within homogeneous sociodemographic cells. The purpose of this adjustment was to reduce the potential bias due to nonresponse, under the assumption that within adjustment cells the characteristics of the respondents are similar to those of the nonrespondents;

3. adjustment for noncoverage within sample PSUs. The purpose of this adjustment was to reduce the potential bias due to the exclusion of BGs/EDs with few Hispanic residents; and

4. poststratified ratio adjustment by age and sex to make the final sample estimates of the population correspond to the most current Bureau of the Census estimates of the civilian noninstitutionlized target population. The ratio adjustment served two purposes. One was to reduce sampling variances, as is normally accomplished by ratio estimates. The second was to dampen any potential biases introduced by the omission of counties with small Hispanic populations.

Noncoverage Analysis

A. Representativeness of the Sampling Frame

Noncoverage of the Hispanic Population in the Southwest has been previously discussed in [3]. As mentioned earlier there were two levels of noncoverage of the Hispanic population in the southwest HHANES as follows:

• noncoverage of the "eligible" Hispanic population residing in excluded counties in the Southwest (about 3 percent noncoverage).
• noncoverage of the "eligible" Hispanic population residing in excluded BGs or EDS (with small numbers of eligible Hispanics) within sample PSUs.

Given the above two levels of noncoverage, the question that arose was "How representative is

the defined HHANES sampling frame of the Mexican origin population living in the Southwest?" It seemed reasonable to believe that the defined HHANES universe was representative of the Southwest since the counties that were included in the frame accounted for approximately 97% of the 1980 Mexican origin population in the Southwest.

In addition, in order to reduce screening costs even further, there was also noncoverage of the Hispanic population at the PSU level since low Hispanic density BGs/EDs were excluded from the sampling frame within each sample PSU. The within sample PSU noncoverage of the Mexican origin population was usually less than 10% and averaged about 7% over all sample PSUs. Although the number of Hispanics omitted was fairly small, an important concern was that the low Hispanic density BGs/EDs contained a disporportinate percentage of high income Hispanic households. It seemed likely that as Hispanics (as other ethnic groups) climb the socioeconomic ladder they are more lilely to move out of their high ethnic concentration areas and assimilate more into the general populations, in which case the sampling frame would underrepresent high income Hispanic households.

In order to investigage the magnitude of the undercoverage of the high income Hispanic households, a comparison was made of the 1979 percent distribution of Hispanic family income for all counties, in-scope counties, and out-of-scope counties in the Southwest. The comparison showed very little difference in the three percent distributions at the Southwest level. A similar comparison was done at the sample PSU level. That is, a comparison was made of the 1979 percent distribution of Hispanic familay income for all BGS, in-scope BGs, and out-of-scope BGS within each sample PSU. The PSU level analysis showed that there were differences between the three percent distributions for certain sample PSUs especially for percent distributions of the in-scope BGs and out-of-scope BGS by income. Therefore it was deemed appropriate to make a noncoverage adjustment within each sample PSU by income. The noncoverage adjustment was computed by taking the ratio of the total Spanish origin families in each sample PSU-income cell to the number of Spanish origin families in in-scope BGs/EDs in the same cell. Table 1 shows the distribution of noncoverage adjustment factors by income for each sample PSU in the Southwest.

As shown in Table 1, the sample PSU level noncoverage adjustments are inflation factors which ranged from a low of 1.00 to a maximum of 1.86. In general there was a direct relationship between income level and size of adjustment factors. That is, as income level increased, their was an increase in the noncoverage and thus an increase in the noncoverage adjustments. Undercoverage of high income households was more prevalent in California than in Texas. Also, there was significant undercoverage in the sample PSUs in Arizona, Colorado, and New Mexico. These adjustments were incorporated as factors in the Southwest HHANES estimator in order to compensate for this undercoverage. By comparison for noncoverage at the Southwest level, corresponding adjustments were computed by income level and

they were equal to 1.04 across the board. Since this adjustment was small, it was decided not to incorporate this noncoverage factor at the Southwest level. However, poststratification by age and sex was incorporated into the estimator and this adjustment probably compensated for the slight undercoverage of high income Hispanic households at the Southwest level.

B. Representativeness of the Sample

Another question that arises once we have selected our sample of elementary units is "How representative is our sample of the population of inference?" In our particular case, how representative is our sample of the Mexican origin population residing in the Southwest? Since we have certain sociodemographic data for both the HHANES sample and for the entire Southwest Mexican origin population based on 1983 Current Population Survey (CPS), we can make direct comparisons of percent distributions for corresponding variables. We can compare CPS percent distributions with weighted or unweighted sample percent distributions by age and sex. Table 2 compares CPS distribution with weighted and unweighted distributions for sample persons (SP's), interviewed SP's, and examined SP's. As can be seen in Table 2, the weighted distributions for SP's and interviewed SP's coincide exactly with the CPS distribution as they should since population estimates were poststratified by age and sex to CPS Southwest totals. The weighted examined percent distribution was also poststratified, however, there are slight differences in the distributions, maybe due to rounding error. It is useful to examine the corresponding unweighted percent distributions in Table 2 and observe the effect of the adjustments due to differential probability weighting, poststratification, noncoverage, interview nonresponse, and examination nonresponse. Table 2 shows that the unweighted sample percent distributions consistently overrepresented the younger and older age groups while underrepresenting the middle age groups. As shown in Table 2, the three unweighted percent distributions are somewhat dissimilar for males, but Table 2 shows that these three distributions are almost identical for females.

Tables 3 compares CPS with weighted and unweighted percent distributions for interviewed SP's by age, sex, and educational level. Table 3 shows that the unweighted distributions are quite different from the CPS distribution by educational level. However, the adjustments made to the sample data bring the weighted distributions into closer alignment with the CPS distribution, especially for the high school (HS) and college distributions. The distributions for each sex indicate the same pattern (Tables not shown).

Table 4 below compares CPS, weighted and unweighted percent distributions by marital status for both sexes, 20-74 years old. As table 4 shows, the weighted and unweighted distributions are quite similar. However, the weighted distribution is more in line with the CPS distribution, although the weighted proportion of married (spouse present) appears to

be overestimated while the proportion of never married appears to be underestimated.

Nonresponse Analysis

Although there are various levels of nonresponse inherent in a multistage sample survey such as the HHANES, two major components of nonresponse will be investigated as possible sources of bias in HHANES estimates.

After investigating the representativeness of the sampling frame as well as the representativeness of the sample itself, a third question arises. Given that adjustments were made to compensate for interview and examination nonresponse (under the assumption that respondents are similar to nonrespondents), "How alike are respondents and nonrespondents in fact?"

A. Interview Nonresponse

In order to compare the characteristics of those sample persons that were interviewed vs those who were not interviewed, sociodemographic data were extracted from the Family Questionnaire for both interviewed and noninterviewed SP's. Comparisons were restricted to sociodemographic characteristics since health data were not available for both interviewed and noninterviewed SP's. Also comparative percent distributions were computed using only unweighted data since sampling weights do not exist for the noninterviewed. A comparison (Tables not shown) of age distributions for interviewed and noninterviewed SP's indicates that the younger age groups are underrepresented while the older age groups are overrepresented.

Table 5 compares percent distributions for interviewed vs noninterviewed by education and sex. It also shows that interviewed and noninterviewed males are very similar by educational level, while interviewed and noninterviewed females are dissimilar especially at the elementary and high school levels.

Table 6 shows that there are large differences between interviewed and noninterviewed SP's by sex and family income especially for the low income group (<$10,000) and the high income group (>$20,000).

Table 7 shows that there are differences (interviewed vs noninterviewed SP's) by sex and household size, especially for 1-2 person households.

B. Examination Nonresponse

The second level of nonresponse that will be investigated is examination nonresponse, that is, a comparison will be made of those SP's who were interviewed and examined vs those who were interviewed and nonexamined. Table 8 compares Census, weighted, and unweighted percent distributions for both sexes examined vs nonexamined by education. As Table 8 shows, the unweighted percent distributions for examined vs nonexamined are very comparable, as are the corresponding weighted distributions.

Table 9 compares percent distributions for examined vs nonexamined by sex and certain health characteristics. With the exception of three health variables, namely, "Glasses or contacts?", "Taking medication for high blood pressure?", and

"Smoking now?", examined SP's are very similar in health characteristics to nonexamined SP's.

Conclusion

This paper has explored noncoverage and nonresponse as possible sources of bias in HHANES estimates. The representativeness of the sampling frame, representativeness of the sample, a comparison of interviewed vs noninterviewed SP's, and comparison of examined vs nonexamined PS's were studied. Although there may still be some bias due to noncoverage, an adjustment by PSU and income level was incorporated into the estimation procedure in order to compensate for this observed bias.

There may be some slight differences in sociodemographic and health characteristics for interviewed vs noninterviewed SP's as well as for examined vs nonexamined SP's. However, the underlying assumption that respondent and nonrespondents are alike within certain adjustment cells appears to be generally valid. Therefore, the HHANES data user can feel confident that valid estimates (with perhaps negligible biases) can be produced and analyzed from this rich health data resource.

References

1. National Center for Health Statistics: Plan and Operation of the Hispanic Health and Nutrition Examination Survey, 1982-84. Vital and Health Statistics. DHHS Pub. No. (PHS) 85-1321. Series 1, No. 19. Public Health Service. Washington, D.C., U.S. Government Printing Office, September 1985.
2. Gonzalez, Joe F., White, Andrew A., and Ezzati, Trena: Sample Design for the Hispanic Health and Nutrition Examination Survey, 1982-84. Proceedings of the Survey Research Methods Section, American Statistical Association, August, 1984.
3. Gonzalez, Joe F., Ezzati, Trena, and White, Andrew A.: Sample Design and Estimation Issues in the Hispanic Health and Nutrition Examination Survey, 1982-84. Invited papers to the 1984 National Center for Health Statistics Data Use Conference on Small Area Statistics, Snowbird, Utah, pp. 18-23.
4. National Center for Health Statistics: Plan and Operation of the Health and Nutrition Examination Survey, United States, 1971-73. By H. W. Miller. Vital and Health Statistics. DHEW Pub. No. (PHS) 79-1310. Series 1, No. 10A and 10B. Public Health Service. Washington, D.C., U.S. Government Printing Office, December 1978.
5. National Center for Health Statistics: Plan and Operation of the Second National Health and Nutrition Examination Survey, 1976-1980. By A. McDowell, A. Engel, J. T. Massey, and K. Maurer. Vital and Health Statistics. DHHS Pub. No. (PHS) 81-1317. Series 1, No. 15. Public Health Service. Washington, D.C., U. S. Government Printing Office, July, 1981.

6. Massey, J. T. and Ezzati, T. M.: HHANES Universe and Estimated Coverage. Unpublished Memorandum. National Center for Health Statistics, August 1981.

7. Gonzalez, Joe F., Ezzati, Trena M., Lago, Josefina, and Waksberg, Joseph: Estimation in the Southwest Component of the Hispanic Health and Nutrition Examination Survey. Proceedings of the Survey Research Methods Section, American Statistical Association, August, 1985.

TABLE 1. SAMPLE PSU LEVEL NONCOVERAGE ADJUSTMENT FACTORS BY INCOME

Income Level	Bee TX	Bexar TX	Cameron TX	Harris TX	Alameda CA	Contra Costa CA	Santa Clara CA	San Diego CA	Los Angeles CA	Pima AZ
<$5,000	1.01	1.01	1.00	1.08	1.10	1.09	1.05	1.07	1.05	1.10
5,000-9,999	1.05	1.01	1.01	1.07	1.13	1.09	1.05	1.08	1.05	1.08
10,000-14,999	1.00	1.01	1.01	1.09	1.12	1.10	1.06	1.09	1.05	1.11
15,000-19,999	1.02	1.01	1.01	1.09	1.11	1.17	1.07	1.10	1.06	1.12
20,000-24,999	1.04	1.01	1.01	1.10	1.13	1.14	1.07	1.13	1.08	1.09
25,000-34,999	1.04	1.02	1.01	1.11	1.13	1.17	1.11	1.14	1.12	1.13
35,000-49,999	1.00	1.03	1.02	1.18	1.13	1.21	1.17	1.22	1.18	1.20
50,000+	1.00	1.07	1.01	1.33	1.18	1.53	1.27	1.44	1.31	1.49

Income Level	El Paso TX	Midland TX	Weld CO	Quay NM
<$5,000	1.00	1.14	1.09	1.01
5,000-9,999	1.00	1.20	1.12	1.15
10,000-14,999	1.00	1.17	1.10	1.14
15,000-19,999	1.00	1.09	1.13	1.01
20,000-24,999	1.00	1.07	1.16	1.04
25,000-34,999	1.01	1.17	1.22	1.17*
35,000+	1.00	1.86	1.42	-

*This adjustment factor applies to $25,000+ income level.

TABLE 2 COMPARISON OF PERCENT DISTRIBUTIONS FOR SAMPLED (SP'S), INTERVIEWED, AND EXAMINED MEXICAN AMERICANS IN THE SOUTHWEST BY AGE AND SEX, 1983

		WEIGHTED			UNWEIGHTED		
CHARACTERISTIC	CENSUS	SP'S	INTER- VIEWED	EXAMINED	SP'S N=4589	INTER- VIEWED N=3926	EXAMINED N=3385
MALES							
TOTAL, 6mo-74yr	100.0	100.0	100.0	100.0	100.0	100.0	100.0
6mo-4yr	11.4	11.4	11.4	12.2	13.5	14.7	15.5
5- 9yr	10.9	10.9	10.9	12.0	13.9	14.9	16.1
10-13yr	8.5	8.5	8.5	9.3	10.6	11.4	12.3
14-19yr	12.5	12.5	12.5	12.7	13.1	13.3	13.0
20-24yr	11.8	11.8	11.8	10.7	7.5	7.3	6.5
25-34yr	19.3	19.3	19.3	18.2	14.0	14.0	12.9
35-44yr	11.0	11.0	11.0	10.7	8.3	7.7	7.4
45-54yr	6.9	6.9	6.9	6.8	9.6	8.2	8.0
55-64yr	4.8	4.8	4.8	4.8	6.8	5.9	5.8
65-74yr	2.8	2.8	2.8	2.6	2.8	2.6	2.5
FEMALES					N=4866	N=4296	N=3812
TOTAL, 6mo-74yr	100.0	100.0	100.0	100.0	100.0	100.0	100.0
6mo-4yr	10.8	10.8	10.8	11.0	12.6	13.0	13.2
5- 9yr	11.2	11.2	11.2	11.6	13.4	13.9	14.6
10-13yr	9.0	9.0	9.0	9.4	9.7	10.2	10.6
14-19yr	12.1	12.1	12.1	11.7	13.2	13.2	12.8
20-24yr	10.6	10.6	10.6	10.6	7.5	7.3	7.3
25-34yr	18.3	18.3	18.3	18.4	14.0	14.1	14.2
35-44yr	11.1	11.1	11.1	11.2	8.6	8.9	9.0
45-54yr	8.0	8.0	8.0	7.7	10.7	9.8	9.5
55-64yr	5.7	5.7	5.7	5.2	6.9	6.4	5.9
65-74yr	3.3	3.3	3.3	3.1	3.4	3.4	3.1

TABLE 3   COMPARISON OF CPS, WEIGHTED, AND UNWEIGHTED PERCENT DISTRIBUTIONS BY AGE, SEX, AND EDUCATION

Characteristic	CPS			Interviewed Weighted			Interviewed Unweighted		
	ELEM	HS	COLLEGE	ELEM	HS	COLLEGE	ELEM	HS	COLLEGE
Both Sexes Total 15-74 yr	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
15-19yr	9.5	23.3	2.5	5.4	24.5	5.8	6.3	30.3	8.6
20-24yr	9.0	21.4	22.3	9.3	19.9	21.5	6.8	15.0	18.1
25-29yr	12.0	15.1	23.6	11.9	15.4	20.0	9.9	13.1	18.8
30-34yr	10.6	13.2	18.5	12.1	11.7	22.1	10.0	10.2	20.7
35-44yr	17.6	13.8	18.7	19.3	14.1	20.2	16.0	12.2	18.2
45-64yr	27.6	11.2	13.0	31.2	12.9	10.0	39.9	17.5	15.1
65-74yr	13.7	2.0	1.5	10.9	1.5	0.4	11.1	1.8	0.5

TABLE 4.   COMPARISON OF CENSUS, WEIGHTED, AND
UNWEIGHTED PERCENT DISTRIBUTIONS BY MARITAL STATUS

Marital Status	Census	Weighted	Unweighted
Total, 20-74 yr Both Sexes	100.0	100.00	100.00
Married -			
Spouse Present	64.8	70.4	71.0
Spouse Absent	5.2	6.0	6.4
Widowed	4.4	3.8	4.8
Divorced	5.7	5.8	5.9
Never Married	19.9	14.1	11.9

TABLE 5   COMPARISON OF INTERVIEWED AND NONINTERVIEWED SP'S BY SEX AND EDUCATION

Sex	Total		Interviewed (Unweighted)			Noninterviewed		
	Interviewed	Noninterviewed	ELEM	HS	COLLEGE	ELEM	HS	COLLEGE
Both Sexes (15-74 yr)	100.0	100.0	37.7	48.4	13.9	40.0	46.5	13.5
Males	100.0	100.0	36.5	47.8	15.7	36.4	49.7	13.9
Females	100.0	100.0	38.7	48.9	12.4	51.0	36.7	12.2

TABLE 6   COMPARISON OF INTERVIEWED AND NONINTERVIEWED SP'S BY SEX AND FAMILY INCOME

Sex	Total		Interviewed (Unweighted)			Noninterviewed		
	Interviewed	Noninterviewed	<$10,000	$10,000-19,999	>$20,000	<$10,000	$10,000-19,999	>$20,000
Both Sexes (15-74 yr)	100.0	100.0	27.7	32.3	40.0	19.0	29.5	51.5
Males	100.0	100.0	23.8	33.2	43.0	16.6	29.8	53.6
Females	100.0	100.0	31.0	31.5	37.5	26.5	28.6	44.9

TABLE 7   COMPARISON OF INTERVIEWED AND NONINTERVIEWED SP'S BY SEX AND HOUSEHOLD SIZE

Sex	Total		Interviewed (Unweighted)			Noninterviewed		
	Interviewed	Noninterviewed	1 - 2	3 - 4	5+	1 - 2	3 - 4	5+
Both Sexes (15-74 yr)	100.0	100.0	19.4	37.8	42.8	9.5	40.5	50.0
Males	100.0	100.0	19.0	37.1	43.9	7.3	41.1	51.7
Females	100.0	100.0	19.8	38.4	41.8	16.3	38.8	44.9

TABLE 8 COMPARISON OF CENSUS, WEIGHTED, AND UNWEIGHTED PERCENT DISTRIBUTIONS FOR
EXAMINED VS NONEXAMINED MEXICAN AMERICANS BY EDUCATION

Education	Census	Weighted Examined	Weighted Nonexamined	Unweighted Examined	Unweighted Nonexamined
Total, Both Sexes (20-74 yr)	100.0	100.0	100.0	100.0	100.0
Elementary	42.5	39.9	37.4	44.8	42.4
High School	42.2	43.1	47.5	40.5	44.8
College	15.3	17.0	15.1	14.7	12.8

TABLE 9 COMPARISON OF WEIGHTED PERCENT DISTRIBUTIONS FOR EXAMINED AND
NONEXAMINED MEXICAN AMERICANS BY SEX AND CERTAIN HEALTH CHARACTERISTICS

Variable	Males Examined	Males Nonexamined	Females Examined	Females Nonexamined
Last Time to M.D.				
Total	100.0	100.0	100.0	100.0
<1 month	17.3	13.9	24.9	31.6
1-5 mos.	27.0	27.0	38.5	34.9
6-11 mos.	20.5	15.7	17.4	14.7
1-4 yrs.	29.6	38.9	17.7	16.5
5 yrs or more	5.4	5.0	1.3	2.4
Don't know	0.1	0.5	0.2	0.0
Glasses or Contacts?				
Yes	74.5	68.0	77.5	79.8
No	25.5	33.0	22.5	20.2
Taking Medication For High Blood Pressure?				
Yes	45.2	62.1	55.7	75.8
No	54.8	37.9	44.3	24.2
Smoke Now?				
Yes	62.4	73.7	65.3	75.1
No	37.6	26.3	34.7	24.9
Ever Told Rheumatic Fever?				
Yes	1.3	2.0	1.6	1.0
No	98.7	98.0	98.4	99.0
Rheumatic Heart Disease				
Yes	0.2	0.0	0.7	1.0
No	99.8	100.0	99.3	99.0
Heart Murmur?				
Yes	3.1	3.1	4.3	4.0
5 yrs or more	96.9	96.9	95.7	96.0
Heart Failure?				
Yes	0.9	0.2	0.9	0.3
No	99.1	99.8	99.1	99.7
Kidney Problems?				
Yes	7.0	6.6	16.4	16.0
No	93.0	98.4	83.6	84.0
Glaucoma?				
Yes	0.4	0.0	0.8	1.4
No	99.6	100.0	99.2	98.6
Cataracts?				
Yes	3.4	2.7	3.0	5.5
No	96.6	97.3	97.0	94.5