

SAMPLE DESIGN FOR THE HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY, 1982-84

Joe Fred Gonzalez, Jr., Andrew A. White, and Trena Ezzati
National Center for Health Statistics

INTRODUCTION

The Hispanic Health and Nutrition Examination Survey (HHANES) is the first large-scale survey designed to assess the health and nutritional status of the Hispanic population living in selected areas of the United States. The general structure of the HHANES sample design and operation was similar to that of the previous three National Health Examination Surveys,¹⁻⁴ the first National Health and Nutrition Examination Survey (NHANES I)⁵ and the second National Health and Nutrition Examination Survey (NHANES II)⁶ conducted by the National Center for Health Statistics (NCHS). The major difference between the HHANES and these previous NCHS surveys was that the HHANES was a survey of a special subgroup of the U.S. population in selected areas of the U.S. rather than a national probability sample.

Persons of Hispanic origin comprised an estimated 6.4 percent of the total 1980 United States population.⁷ Health data have been collected for Hispanics in previous NCHS surveys; however, Hispanics were sampled according to their proportion of the total population. Because the proportion of Hispanics in the United States population is relatively small, the number of Hispanics included in these surveys has generally been inadequate for making precise estimates of their health parameters in sufficient detail to be useful. Previous national surveys have also been unable to divide "Hispanics" into smaller, more homogeneous population groups for analysis. The term "Hispanic" is one of convenience and is used to describe collectively United States residents of Mexican, Puerto Rican, Cuban, Central American, South American, and other Spanish backgrounds. Although much of the Hispanic population is concentrated in high density urban areas, a sizable proportion of this population is dispersed throughout the general population. However, the three major Hispanic subgroups, i.e., persons of Mexican, Puerto Rican, and Cuban descent are highly concentrated in the five Southwest States (Arizona, California, Colorado, New Mexico, and Texas), in the New York City area, and Dade County (Miami), Florida, respectively. Due to budgetary, operational, and data analysis considerations, the NCHS decided that the most efficient and cost effective method of surveying Hispanics would be to survey the three major Hispanic subgroups in their respective areas of concentration. As a result, valid statistical inferences can be made only to the Mexican-American population residing in the Southwest, the Puerto Rican population residing in the New York City area, and the Cuban-American population residing in Dade County (Miami), Florida. With a probability sample large enough to provide sufficient numbers of these Hispanic subgroups, the HHANES can provide important information on the health and nutritional status of these Hispanic Americans.

The primary objectives of the HHANES were to

provide prevalence estimates for certain disease conditions and to provide normative data on various health and nutritional characteristics for the target population. It was also desirable to investigate familial aggregation of disease conditions, nutritional status, and patterns of medical care utilization and expenditures. The HHANES added a new dimension to NCHS survey planning because it was the first such survey to be designed with a family analysis unit in mind.

For the HHANES, a complex, multistage, stratified, probability cluster design was used to survey "eligible" Hispanics in a limited number of selected areas of the United States. The HHANES was a stratified four stage cluster sample selected with probabilities proportional to size (PPS) of the Hispanic population at each stage. 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.

This paper describes the definition of the universes, construction of the sampling frames of PSUs, the stratification and selection methods that were used at each stage of the multistage probability design, and the basic framework for constructing estimates for the survey.

DEFINITION OF THE UNIVERSE, PRIMARY SAMPLING UNITS, AND CONSTRUCTION OF THE SAMPLING FRAME

Ideally, the target population for the HHANES should be all households with at least one member of Hispanic origin. However, the U.S. includes States and counties whose Hispanic population and/or density is very small, and if sampled, the costs of screening enough households to locate a reasonable sample size in these areas would be quite high. This screening cost would significantly reduce the resources available for examination. The loss of information from the reduced sample size is believed greater than the advantages resulting from the somewhat greater representativeness of the sample, if the entire U.S. were sampled.

Consequently, the universe for the HHANES was restricted to counties in the three areas of the United States which had a sufficient number and/or percentage of Hispanics to render it economically feasible to establish and operate an examination center over a four to seven week time period.

Using 1980 Census data, the 3,137 counties in the United States were distributed into cells according to the percent and number of Hispanics in the county.⁸ Those counties or groups of contiguous counties which contained a small number or proportion of Hispanics were excluded from the universe.

Initially, 252 of the total 3,137 counties were identified for inclusion in the universe if they satisfied one of the four following criteria based upon a combination of the

county's total number of Hispanics and the proportion of the total 1980 county population that was of Hispanic (Mexican, Puerto Rican, Cuban and "other" Hispanic) origin:

1. The county Hispanic origin population was greater than 30,000.
2. The county Hispanic origin population was greater than or equal to 10,000 and less than 30,000 and it comprised at least five percent of the total county population.
3. The county Hispanic origin population was greater than or equal to 5,000 and less than 10,000 and it comprised at least 10 percent of the total county population.
4. The county Hispanic origin population was less than 5,000 and it comprised at least 15 percent of the total county population.

Secondly, 23 of the 252 counties included by one of the four initial criteria were excluded from the universe for economic and logistical considerations because they were not located in one of the five Southwestern States (Arizona, California, Colorado, New Mexico, and Texas), the New York City area, or Miami, Florida area.

The resulting 229 counties were grouped into 210 PSUs, defined as a single county or a small group of counties with a 1980 Hispanic population of at least 1,000. Finally, the 210 PSUs were grouped to form three separate universes: HHANES Mexican Universe for the Southwest, 193 PSUs; HHANES Cuban Universe for Dade County (Miami, Fla.), 1 PSU; and the HHANES Puerto Rican Universe for the New York City area, 16 PSUs. These three universes of PSUs include approximately 76 percent of the 1980 Hispanic origin population in the United States.

The HHANES Mexican origin universe of first stage units (193 PSUs) 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. As a means of reducing screening costs (the screening procedure will be described later) in low Hispanic density areas, a small percentage (usually less than 10 percent) of the Mexican origin population within each sample PSU was not covered because block groups (BGs) or enumeration districts (EDs) that did not contain a minimum number (between 50 and 100) of "eligible" Hispanics were excluded. The count of "eligible" Hispanics within a given BG/ED was defined as the number of Mexican origin persons plus a certain (PSU-specific) percentage of persons of "other Spanish" origin who were assumed to be of Mexican origin (this will be discussed in more detail in the section on within-PSU design). Hence, the net coverage rate of the 1980 Mexican origin population in the Southwest was approximately 87 percent (.97 X .90).

The HHANES Cuban origin universe, which consisted of only one PSU, Dade County (Miami, Fla.), included about 57 percent of the 1980 Cuban origin population in the United States. Because low Hispanic density BGs within Dade County were excluded, the Cuban universe covered about 96 percent of the Cuban origin population in Dade County, Florida. Thus, the net coverage

of the Cuban origin population in the United States was approximately 55 percent (.57 X .96).^a

The HHANES Puerto Rican origin universe of first stage units (16 PSUs) included about 59 percent of the 1980 Puerto Rican origin population in the United States and about 90 percent of the 1980 Puerto Rican origin population in the combined States of Connecticut, New Jersey, and New York. As in the Southwest and Dade County Universes, BGs with low Hispanic density were excluded from within the sample PSUs in the Puerto Rican universe. Thus, the net coverage of the 1980 Puerto Rican origin population in the combined States of Connecticut, New Jersey, and New York was approximately 81 percent (.90 X .90).

Since three separate universes were created, statistical inferences will be made: for the Mexican origin population residing in the defined universe for the five southwestern States; for the Puerto Rican origin population residing in the defined universe for the New York City area; and for the Cuban origin population residing in Dade County (Miami), Florida.

STRATIFICATION OF PRIMARY SAMPLING UNITS⁹

Stratification of units prior to sample selection is a technique which is widely used in scientific sampling. The goal of stratification is to reduce the variance of the survey estimates by forming strata which are composed of homogeneous units. For a multistage design this can be accomplished in part by partitioning the universe of PSUs into optimally homogeneous strata. Based on sampling theory, PSUs should be stratified according to the survey variables of interest, however, as is often the case, estimates of survey variables are generally unknown prior to the conduct of a particular survey. Therefore, stratification is usually based on the sociodemographic characteristics of the PSUs which are believed to be correlated with the survey variables.

Since 1980 Census information for the Mexican origin population in the Southwest PSUs was unavailable prior to stratification, information based on the number of all Hispanics was used to compute the measures of size (MOS) of PSUs and to stratify the PSUs. The five Southwest PSU characteristics that were used as stratification variables were:

- o number of Hispanics
- o percent Hispanic
- o ratio of the 1980 to the 1970 Hispanic population
- o median income
- o percent urban

For the New York City area component of the HHANES, the MOS for the PSUs was the number of Puerto Ricans. Therefore, the corresponding stratification variables were in terms of the number of Puerto Ricans in the New York City area.

The PSUs from the Southwest and the New York City area universes were stratified in a similar fashion. Since the Miami area component of the HHANES consisted of only one PSU, Dade County, stratification of counties was not required there. The methodology for stratifying the first-stage units for the other two areas is described as follows.

A critical sample design feature for the HHANES was that the Southwest strata be of equal Hispanic population size while the New York City area strata be of equal Puerto Rican population size. This was a design requirement for the HHANES since equal-size strata generally come close to minimizing sampling variances, and at the same time permit the same number of sample interviews and examinations at each survey location. This requirement was satisfied by forming equal size strata and then applying the same sampling fraction to each stratum.

The Statistical Analysis System (SAS) PROC CLUSTER¹⁰ was the technique that was chosen for the HHANES stratification problem. The SAS routine PROC CLUSTER (outlined by JOHNSON¹¹) is a multivariate procedure which uses a hierarchical algorithm for grouping similar vector observations. A major drawback of the algorithm is its inability to impose constraints on the cluster sizes. However, by iteratively applying this SAS procedure, the clustering process was controlled to yield strata of approximately equal size. Fourteen strata (2 certainty, 12 non-certainty) were formed for the Southwest universe and seven strata (3 certainty, 4 non-certainty) were formed for the New York City area universe.

SELECTION OF PRIMARY SAMPLING UNITS

The Southwest and the New York City area HHANES universes of PSUs were stratified according to the five demographic characteristics discussed earlier. Moreover, it was desirable to maximize the probability that each of the five Southwest States in the universe would contribute to the final sample. Therefore, during PSU selection, a slightly modified version of a procedure introduced by Goodman and Kish¹² and summarized in Kish,¹³ was used to obtain a balanced sample with respect to State while retaining a true probability sample design. A detailed description of this controlled selection process and its application to health examination surveys is given in other NCHS reports.^{4,6}

In the following paragraphs, the general process of controlled selection is briefly described and then the modifications used to select the sample for the HHANES in the Southwest are discussed. Controlled selection was not used in the HHANES Puerto Rican and Cuban components.

The goal of controlled selection was to introduce controls beyond stratification into the sampling process. The first step involved defining a set of admissible patterns so that sampling according to any pattern would result in the desired distribution of PSUs across the control classes. A pattern, or potential sample, was admissible if the difference between the number of PSUs in the pattern in each control class was within one of the number expected if a sample of the same size were to be drawn strictly at random from the universe with probability proportional to size. The total set of patterns was formed so that by choosing one pattern from the set, the probability of selecting any given PSU in the universe was proportional to its population. Each pattern within the set was usually assigned a probability of selection based on the accumulated size of the sample PSUs within the pattern. The sum of the probabilities of selection over all patterns was

held equal to one. A pattern was then picked at random.

For the HHANES, having State as the single control variable during selection allowed the procedure to be modified to some extent for simplicity. The goal was to preserve the representation of the five southwest States in the final sample of 14 PSUs (selecting one PSU per stratum) in proportion to their Hispanic population. For the HHANES, 100 equally probable patterns were defined where each pattern detailed for each stratum which State (of those represented in the stratum) should contribute the single PSU to represent that stratum in a potential sample based on that pattern. These patterns were defined so that the number of patterns that would subject any given PSU to inclusion in a final sample was proportional to its Hispanic population. Each of the 100 patterns met the admissibility criteria stated in the general description above.

The procedure used differed from the usual practice in Goodman-Kish controlled selection in that some of the 100 patterns were duplicates. As applied previously, patterns were not usually assigned equal probabilities, and indeed the duplicate patterns could have been collapsed and assigned proportionally higher selection probabilities. Nevertheless by allowing duplicates, pattern construction was made somewhat more straightforward and the same goal was accomplished. One of the 100 admissible patterns was then selected by simple random sampling. The State representation having thus been set for each stratum, the final PSUs were selected using PPS sampling from among the PSUs associated with the pattern for each stratum.

As mentioned earlier, controlled selection of PSUs was only used in the Southwest. For the New York City area component, one PSU per stratum was selected with PPS. For the Miami area component, Dade County was the only PSU selected. Table 1 shows the sample PSUs selected for the three components of the HHANES.

WITHIN PSU DESIGN^D

Definition of the eligible within-PSU population. Within the PSUs selected at the first stage of sampling, the in-scope population consisted of all households and residents of group quarters (noninstitutional) containing one or more "eligible" Hispanics. As mentioned earlier, "eligible" Hispanics within a PSU in the Southwest were defined as residents of Mexican origin plus a percentage of persons of "other Spanish" origin within a BG or ED who were assumed to be of Mexican origin. The rationale for this was that the great majority of persons of "other Spanish" origin in the Southwest consider themselves as "Hispanos" instead of Mexicans. "Hispanos" are descendants of immigrants from Spain who moved into the Southwest before 1850 and thus do not consider themselves "Mexican" but are in-scope for the HHANES. The Ethnic and Spanish Statistics Branch, Bureau of the Census, provided assistance in this matter.

"Eligible" Hispanics for the New York City area component of the HHANES were defined as residents of Puerto Rican origin, while "eligible" Hispanics for the Dade County component of the HHANES were defined as resi-

Table 1 - SAMPLE PSUS FOR THE HISPANIC HANES

Southwest component (County and State)	Total 1980 Hispanic Population	Percent Hispanic	Desired Sample Persons
Los Angeles, CA.*	2,065,727	27.6	2,435
Bexar, TX.*	460,911	46.6	543
Harris, TX.	369,075	15.3	565
El Paso, TX.	297,001	61.9	582
San Diego, CA.	275,176	14.8	662
Cameron, TX.	161,632	77.1	628
Santa Clara, CA.	226,611	17.5	502
Pima, AZ.	111,418	21.0	606
Contra Costa, CA.	55,977	8.5	539
Glasscock- Midland, TX.	12,699	15.1	643
Alameda, CA.	129,962	11.8	577
Weld, CO.	21,017	17.0	641
Quay, NM.	3,753	35.5	536
Bee, TX.	11,914	45.8	541
Total			10,000

New York City Area component (County and State)	Total 1980 Puerto Rican Population	Percent Puerto Rican	Desired Sample Persons
Bronx, NY.*	320,098	27.4	1,062
Kings, NY.*	279,646	12.5	928
New York, NY.*	166,328	11.6	552
Fairfield, CT.	29,527	3.7	338
Hudson, NJ.	55,828	10.0	336
Nassau, NY.	13,984	1.1	361
Queens, NY.	83,425	4.4	336
Total			3,913

Miami Area component (County and State)	Total 1980 Cuban Population	Percent Cuban	Desired Sample Persons
Dade County, FL.*	407,253	25.1	2,266

*Certainty PSU.

dents of Cuban origin.

Because Hispanics constitute a minority of the population in most PSUs, considerable screening of households was required to locate a sample of "eligible" Hispanic households within a PSU. As a means of reducing costs due to the household screening procedures, BGs/EDs with very low "eligible" Hispanic density were excluded from each of the three universes and were considered out-of-scope. The overall goal was to attain a minimum of 90 percent coverage of the eligible Hispanic population within each sample PSU. In addition, certain types of living quarters were considered out-of-scope.

The following were considered to be out-of-scope:

- a. BGs (or EDs where BGs were not defined) with a small number of "eligible" Hispanics. The variable cutoff (minimum required number of "eligible" Hispanics) was between 50 and 100 for BGs or EDs in the Southwest and 100 Cubans for BGs in Dade County, and between 6 and 100

Puerto Ricans in the New York City area.

- b. institutionalized populations,
- c. Indian reservations, and
- d. military installations.

Stratification of BGs and EDs. After eliminating BGs and EDs that were out-of-scope as described above, the remaining BGs and EDs were stratified by two stratifying variables, an economic index and Hispanic density. The economic index was contract rent, in areas in which rental housing dominated, and value of home, in areas of predominantly owned homes. The rationale for sorting the within-PSU file in this manner was to achieve implicit stratification of BGs/EDs when a systematic sample of segments was drawn (to be further discussed in the section on selection of Secondary Sampling Units).

Formation of Secondary Stage Sampling Units (SSU's). In general, the Secondary Sampling Units (SSU's) were segments, which were mainly blocks or combinations of neighboring blocks (generally contiguous) in urban areas. In rural areas the SSU's were similar blocks or other types of area segments. However, in non-block areas where segments were mapped and selected in a manual operation, the selection was a two-step process, i.e., EDs were selected first and then segments were selected within the sampled EDs.

The within PSU sampling procedures were designed to yield an approximate self-weighting sample of households, i.e., every eligible household had about the same probability of selection.

It was a design requirement to sample an average of six Hispanic households per segment. Assuming three sample persons per household,¹⁴ this equated to a desired sample size of 18 sample persons per segment. The measure of size (MOS) for each segment in the Southwest and the New York City area that was established was approximately equal to the sum of

- 3/4 of "eligible" Hispanics aged 6 months - 19 years;
- 1/2 of "eligible" Hispanics aged 20 to 44 years; and
- all of "eligible" Hispanics aged 45 to 74 years.

In the Miami area, the measure of size for each segment was approximately equal to the sum of

- all of "eligible" Hispanics aged 6 months - 19 years;
- 2/3 of "eligible" Hispanics aged 20 -44 years; and
- all of "eligible" Hispanics aged 45 -74 years.

The age-specific sampling rates that were used to compute the MOS of segments were those used to select sample persons within sampled households (to be described in detail later). These age-specific sampling rates were used within the household to ensure sufficient sample sizes in the desired estimation cells. These sampling rates were therefore used to compute the MOS of segments because they reflected the number of expected sample persons somewhat better than total population or total households and thus provided greater control of segment size.

The 1980 Census tapes did not contain exact counts of "eligible" Hispanics in the required age-ranges, but reasonable approximations were made. As indicated above, the number of sample persons desired per segment was 18. However, it was necessary to start off with a larger segment size to allow for: (a) normal nonresponse; and (b) a reserve in case the nonresponse was higher than expected, or there had been important changes in age distributions or the total population of the area since 1980.

From past NHANES' results, the expected combined screener and interview response rate was assumed to be 80 percent. Including an additional reserve for unexpected contingencies increased the 18 sample persons per segment to 27. Therefore, the minimum segment MOS was 27.

Selection of secondary sampling units (SSU's).

The MOS was calculated for each block and ED. If the block MOS was less than 27, the block was combined with as many blocks following on the sorted list as necessary to reach the minimum of 27. However, the combinations were kept within a BG. If the combination of blocks reached the end of the BG without reaching 27, the combination was added to the immediately preceding block(s). The individual blocks, combinations of blocks, and area segments within the EDs were the actual sampling units (segments). In selecting the sample, a skip interval was determined by dividing the total PSU measure of size by the number of segments required to yield the desired number of sample persons from the PSU. Using a random start between 0 and the skip interval, a systematic sample of segments was selected with probability proportional to size (PPS). That is, the probability of selection depended on the MOS for each segment.

Within-segment selection of households and sample persons. After selecting the sample segments, households were listed within each segment. Depending on the MOS of a particular segment, all or a subsample of the listed households were screened to determine whether any persons self-identifying as Hispanic were present. The subsampling rate for selecting the households to be screened was achieved by subsampling at a rate of 27 divided by the measure of size of the segment.

A modified Perkin's Stop Rule was employed in order to control the expected number of sample persons in each PSU. The maximum sample consisted of the "basic" sample of households plus several groups of reserve households (about 20 percent of "basic" sample) to allow for non-response and changes in Hispanic population distribution. Each group of reserve households was a random subsample of households listed within the segments selected in the PSU's. In certain PSUs it was necessary to de-select some sample households due to a larger than expected number of sample persons. However, this was also done on a random basis within the entire PSU to avoid any potential biases in the de-selection process. Once the eligible households were sampled, every family within the household was eligible to participate in the HHANES if it contained at least one "eligible" Hispanic (that is, a person whose self-reported national origin was the Hispanic group "eligible" in that area). Every member 6 months through 74 years of age

(who usually resides at the household) within an eligible family had a probability of selection since persons were subsampled across eligible families at the same age-specific sampling rates used to compute the measure of size of segments.

ESTIMATION PROCEDURES

The estimation procedure was designed to yield statistics that are essentially unbiased. The three basic components normally included in the estimation procedures for previous NCHS surveys are:

1. inflation of sample person observations by the product of the reciprocals of the probabilities of selection at each stage;
2. adjustment for nonresponse within homogeneous sociodemographic cells (e.g., age-sex-income cells). The purpose of this adjustment is to reduce the potential bias due to nonresponse, under the assumption that within cells the characteristics of the respondents are similar to those of the nonrespondents; and
3. poststratified ratio adjustment within certain age-sex demographic cells in order to make the final sample estimates of the population agree exactly with corresponding Bureau of the Census population figures for the same cells. This type of adjustment achieves most of the gains in precision which would have been attained if the sample had been drawn from a population stratified by the same demographic variables.

For the HHANES, noncoverage of some of the "eligible" Hispanic population within excluded counties and BGs and EDs with small numbers of Hispanics was an additional estimation issue which was investigated. A ratio adjustment was made to compensate for this noncoverage and thus reduce the potential bias due to noncoverage. Final estimation procedures will be developed after the survey results have been edited and processed.

The HHANES was designed so that separate estimates could be produced for the three target populations within each of the three sampling frames. Further details of the HHANES estimator may be obtained from the Office of Research and Methodology, NCHS.

Footnotes

^aSince the official date of the 1980 census was April 1, 1980, the Cubans who came to the U.S. in the period immediately following April 1 may not be included in the 1980 Census count. It is generally believed that most of them are in Dade County, and as a result the percentage of Cuban Americans covered in the United States by the survey was probably greater than 55 percent.

^bA major portion of this section was extracted from documentation produced under contract by Westat, Inc., the data collection contractor for the HHANES. Mr. Joseph Waksberg and Ms. Josefina Lago were the project statisticians at Westat.

Acknowledgement

The authors would like to express their gratitude to Mr. Tom Darby, Computer Users

Staff, for doing an excellent job of producing the maps that were used for the paper presentation. Also, the authors appreciate the assistance of Ms. Carol Alexander, Division of Health Examination Statistics, for her excellent typing and preparation of this manuscript.

References

1. National Center for Health Statistics: Plan and initial program of the Health Examination Survey. By W.R. Simmons, et al. Vital and Health Statistics. DHEW Pub. no. (HRA)74-1038. Series 1, No. 4. Public Health Service. Washington, D.C. U.S. Government Printing Office, Nov. 1973.
2. National Center for Health Statistics: Plan, operation, and response results of a program of children's examinations. By A.J. McDowell. Vital and Health Statistics. DHEW Pub. no. (HSM)73-1251. Series 1, No. 5. Public health Service. Washington, D.C. U.S. Government Printing Office, Oct. 1967.
3. National Center for Health Statistics: Plan and operation of a Health Examination Survey of U.S. youths 12-17 years of age. By H.W. Miller. Vital and Health Statistics. DHEW Pub. no. (HRA)75-1018. Series 1, No. 8. Public health Service. Washington, D.C. U.S. Government Printing Office, Sep. 1974.
4. National Center for Health Statistics: Sample design and estimation procedures for a National Health Examination Survey of children. By E.E. Bryant, J.T. Baird, and H.W. Miller. Vital and Health Statistics. DHEW Pub. no. (HSM)72-1005. Series 2, no. 43. Health Services and Mental Health Administration. Washington, D.C. U.S. Government Printing Office. Aug. 1971.
5. National Center for Health Statistics: Plan and operation of the Health and Nutrition Examination Survey, United States, 1971-1973. 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, Dec. 1978.
6. 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.
7. 1980 Census of Population Supplementary Report: Persons of Spanish origin by State: 1980. Bureau of the Census Pub. No. PC 80-S1-7. Washington, D.C. U.S. Government Printing Office. Aug. 1982.
8. Massey J.T. and Ezzati T.M.: HHANES universe and estimated coverage. Unpublished memorandum. National Center for Health Statistics, Aug. 7, 1981.
9. Gonzalez J.F. and White A.A.: Stratifying primary sampling units with the SAS cluster procedure. Proceedings of the Social Statistics Section, American Statistical Association, Aug. 1982.
10. SAS Institute, Inc.: SAS Users' Guide, 1979 edition. Raleigh, N.C.
11. Johnson S.C.: Hierarchical clustering schemes. Psychometrika. XXXII:241-54, 1967.
12. Goodman R., and Kish L.: Controlled selection - a technique in probability sampling. JASA 45:350-72, 1950.
13. Kish L.: Survey Sampling. New York. John Wiley and Sons, Inc., 1965.
14. U.S. Bureau of the Census: Persons of Spanish origin in the United States: March 1979. Current Population Reports. Series P-20, No. 354. Washington, D.C. U.S. Government Printing Office, Oct. 1980.