

The Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

Supporting Statement Part B – Revised

2/8/2008

Contact: Lorraine Silsbee, M.H.S. or Paul Sorlie, Ph. D.

6701 Rockledge Drive MSC 7936

Bethesda, MD 20892

Phone : 301-435-0707

FAX: 301-480-1455

E-mail: silsbeeL@nhlbi.nih.gov

B. Collections of Information Employing Statistical Methods

B.1. Respondent Universe and Sampling Methods

B.1.a. Design Summary

The sampling and recruitment plan for the study is designed to support four analysis objectives. To accomplish these objectives, a representative sample of participants in the target areas at each field center is selected. Methods of sample selection, recruitment, and retention are designed to maximize participation rates, minimize non-response, and minimize attrition during the follow-up period.

First, the study sample supports estimates of the prevalence (or mean values) of baseline risk factors for 1) all Hispanics combined in this study; 2) all study Hispanics by community of residence; 3) all study Hispanics by country of origin; 4) to a limited extent all study Hispanics by community of residence controlling for country of origin; and 5) to a limited extent all study Hispanics by country of origin controlling for community of residence. Secondly, the study sample supports evaluation of the relationships between the various risk factors, demographic factors, and cultural factors collected at baseline. Thirdly, the study sample supports evaluation of factors collected at baseline in relation to the incidence of disease and death that will occur during the follow-up period. Within the current follow-up period, there will be a small number of events (about 100) for broad analysis of risk factors and incidence. In the longer follow-up (proposed but not funded at this time) there will be increasing numbers of events for more detailed and complex analysis. Lastly, the study sample supports the future potential for a re-examination and re-measurement of the same factors collected at the baseline examination. A re-examination of these cohorts is proposed, but not funded at this time, and would provide estimates of factors related to change in the measured characteristics, and would provide the

ability, with further follow-up, to estimate the impact on disease and death.

B.1.b. Statistical Considerations and Power

The following table describes the anticipated sample size at the baseline examination by country of origin and community.

<u>Community</u>	<u>Country of Origin</u>				<u>Total</u>
	<u>Central So.Amer</u>	<u>Cuban</u>	<u>Mexican American</u>	<u>Puerto Rican Dominican</u>	
Bronx	320	40	280	3360	4000
Chicago	440	40	2440	1080	4000
Miami	1400	2400	40	160	4000
San Diego	80	0	3880	40	4000
Total	2240	2480	6640	4640	16000

As can be seen from this table, the sample size for each community is large (4000 each), and it also is large for each country of origin group (excess of 2000 in each group). While specific examples of prevalence and mean value comparison and subsequent statistical power are shown below, the sample sizes in this table are appropriate for comparisons by community, or comparisons by country of origin (sub-objectives 2 and 3 above). For sub-objectives 4 and 5 above, however, analyses are more limited. Because of the geographical clustering by country of origin, there is not complete diversity of origin for all communities. As seen below, groups of 250 or more can produce estimates of mean values with sufficiently high statistical power so that comparisons for some of the community and ethnic group combinations can be made. For example, within the Bronx and Chicago, reasonable comparisons can be made among Central/South Americans, Mexican Americans and Puerto Ricans/Dominicans. Because San Diego is nearly all Mexican American, it is the only community that will not have within-

community comparisons by country of origin.

The second analytical goal is to assess the relationships among the variables collected at baseline. These analyses are nearly unlimited in number, but include comparisons of risk factors and other measured characteristics. A list of the types of analyses would take many pages, but examples of some analyses would include evaluation of the relationship between body mass index and blood pressure or cholesterol; the relationship between hypertension and length of residence in the U.S.; the relationship between dietary factors of fat and carbohydrates with obesity or lipids; the relationship between perceived Hispanic identity and use of health care services; the relationships between hearing loss and occupations that involve machine noises; the relationships between sleep quality and metabolic syndrome; the relationship between daily physical activity and diabetes or obesity. Minimum sample sizes for comparative examples are shown below.

The third analytical goal is to assess the relationship between baseline characteristics and incident disease and death. As indicated, the study is likely to observe around 100 incident heart attacks in the 3 and ½ year follow-up period. This will allow estimation of relative risks for the study combined. In the total study population, for risk factors with a p-value of 30 percent, a relative risk of 2.0 can be detected with 84 percent power and an alpha of 0.05. For more detailed analysis of risk factors and incidence, the study will need to be renewed to obtain additional endpoint events.

The final analytical goal is to assess change in characteristics over a six year period. This would require a new examination of the cohort six years after the first and again requires study renewal.

Statistical power calculations are shown below for a variety of examples. These examples show the minimum sample size necessary for comparisons of mean values or proportions. These comparisons could be within the groups in the table above related to the first goal, or comparisons similar to those in the second goal.

Sample size calculation

Listed below are minimum samples sizes within each group to compare two groups assuming an alpha of 0.05 and a power of 0.80. Standard errors necessary for these calculations were based on data from existing studies. Examples of comparisons are shown, though this is only the briefest list of potential research questions.

1. Mean systolic blood pressure difference of 10mm HG: $n = 64$
Examples include comparing obese and non obese; comparing by country of origin within Chicago; comparing those born the U.S. vs immigrants.
2. Mean total plasma cholesterol difference of 20 mg/dl: $n = 64$
Examples include comparing those who eat traditional Latino foods vs those who eat American foods; those with high activity levels vs those with low levels.
3. Mean difference in BMI of 5 units: $n = 17$
Examples include comparison of recent immigrants vs long term immigrants; those who identify with Latino culture and traditions vs those who don't.
4. Mean difference in glucose measured by the oral glucose tolerance test of 20 mg/dl: $n = 37$
Examples include comparison of those with and without sleep disturbances; those smoking vs non-smokers.
5. Difference in prevalence of diabetes (10% as compared to 15%): $n = 721$
Examples include comparison of obese vs non-obese persons; comparison of Central Americans vs Cubans in Miami.
6. Difference in prevalence of hypertension (30% as compared to 20%): $n = 376$
Examples include comparison of persons of high vs low incomes; comparison of those with visceral adiposity (waist girth) vs those without.

7. Difference in prevalence of reported CHD (5% as compared to 10%): n = 464

Examples include comparison of persons with supportive social networks compared to those without; comparison of those in the upper quartiles of C-reactive protein (an inflammatory marker) and those in the lower quartile.

8. Difference in smoking prevalence (40% compared to 50%): n = 407

Examples include comparison of persons of Mexican origin who live in Chicago and San Diego; comparison of those in the highest quartile of the ankle-brachial index compared to those in the lowest quartile.

9. Difference in prevalence of hearing disorders (20% compared to 40%): n = 91

Examples include comparison of persons with occupations with noise exposure vs those without; comparison of persons in the highest quartile of cognitive function with those in the lowest quartile.

10. Difference in prevalence of any tooth loss (25% compared to 35%): n = 348

Examples include comparison of persons born in the US vs foreign born; comparison in those with high education vs low education.

B.1.c. Sample Selection

Sample selection is accomplished through a two-stage area probability sample implemented for each site. At the first stage, a stratified sample of Census block groups is selected. Stratification factors common across the four field centers are (1) low versus high SES (as measured by the proportion of persons with at least a high school education) and (2) low vs. high concentration of Hispanic/Latino households, resulting in four strata per field center. Selection of block groups is carried out proportionately with respect to the SES strata and disproportionately with respect to the Hispanic/Latino concentration strata, that is, block groups in the high concentration stratum are selected at a higher rate than those in the lower concentration stratum. This over-sampling is carried out to maximize efficiencies in the field by increasing the probability that a selected household is a Hispanic/Latino household. In addition to these four strata, block groups in the Coop City area are isolated into a 5th stratum in the Bronx, and block groups representing high concentration areas for Central and South Americans

are isolated into a 5th stratum in Miami. Both of these ‘special’ strata are defined to ensure selection of adequate numbers of households in the respective areas.

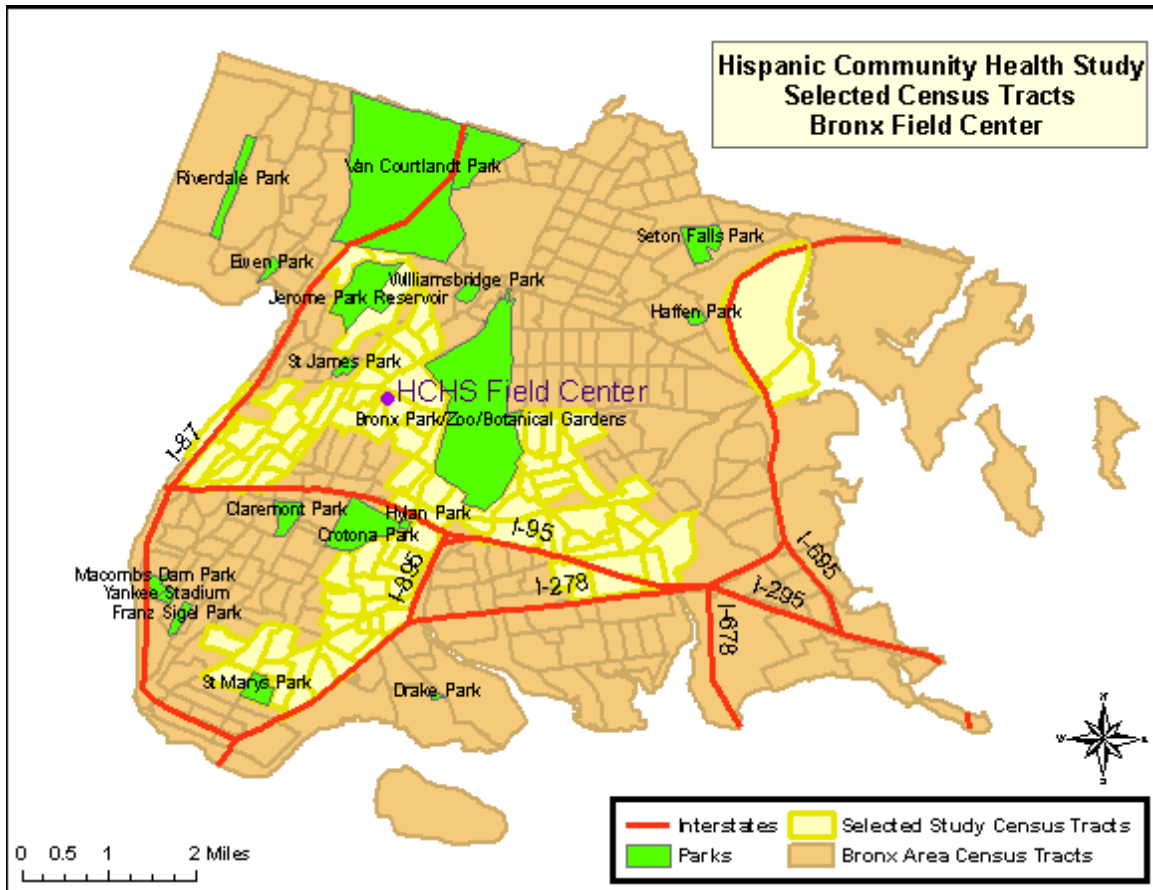
At the second stage, households in the sampled block groups are selected from a dual frame constructed from non-overlapping lists of postal addresses and Hispanic/Latino surnames. Addresses are selected from the surname list at a higher rate than from the postal list, to further maximize efficiency of field operations by increasing the probability that a selected household is a Hispanic/Latino household. Selected households are screened for eligibility, where eligibility is defined as at least one Hispanic/Latino household member aged 18-74 years. Eligible households in which all Hispanic/Latinos in the target age range are at least 45 years of age are selected with certainty (probability of selection = 1), while all other households are selected with probability ($0 \leq p < 1$) based on the expected household composition for the area. Once a household is selected, all members of the household are invited to participate. This household selection algorithm is designed to provide the target age distribution for the HCHS/SOL study, namely, 62.5% of participants aged 45-74 years and 38.5% aged 18-44 years, and to minimize the amount of information required for screened households that may not be selected for participation. Selection of households corresponds to an over-sampling of Hispanic/Latinos in the older age range, which is necessary given the age distribution of Hispanic/Latinos currently living in the US.

Recruitment is planned for a three-year period. The sample of households in each target area is randomly allocated to each of the three years of recruitment. Within each recruitment year, we anticipate fielding the sample in waves, with each wave corresponding to a random sub-sample of the original sample of households allocated to that year.

B.1.d. Four thousand (4,000) persons aged 18-74, who self-identify as Hispanic/Latino origin, independent of country of origin, will be selected from each of four separate communities:

Bronx: over 644,000 residents of Hispanic/Latino origin

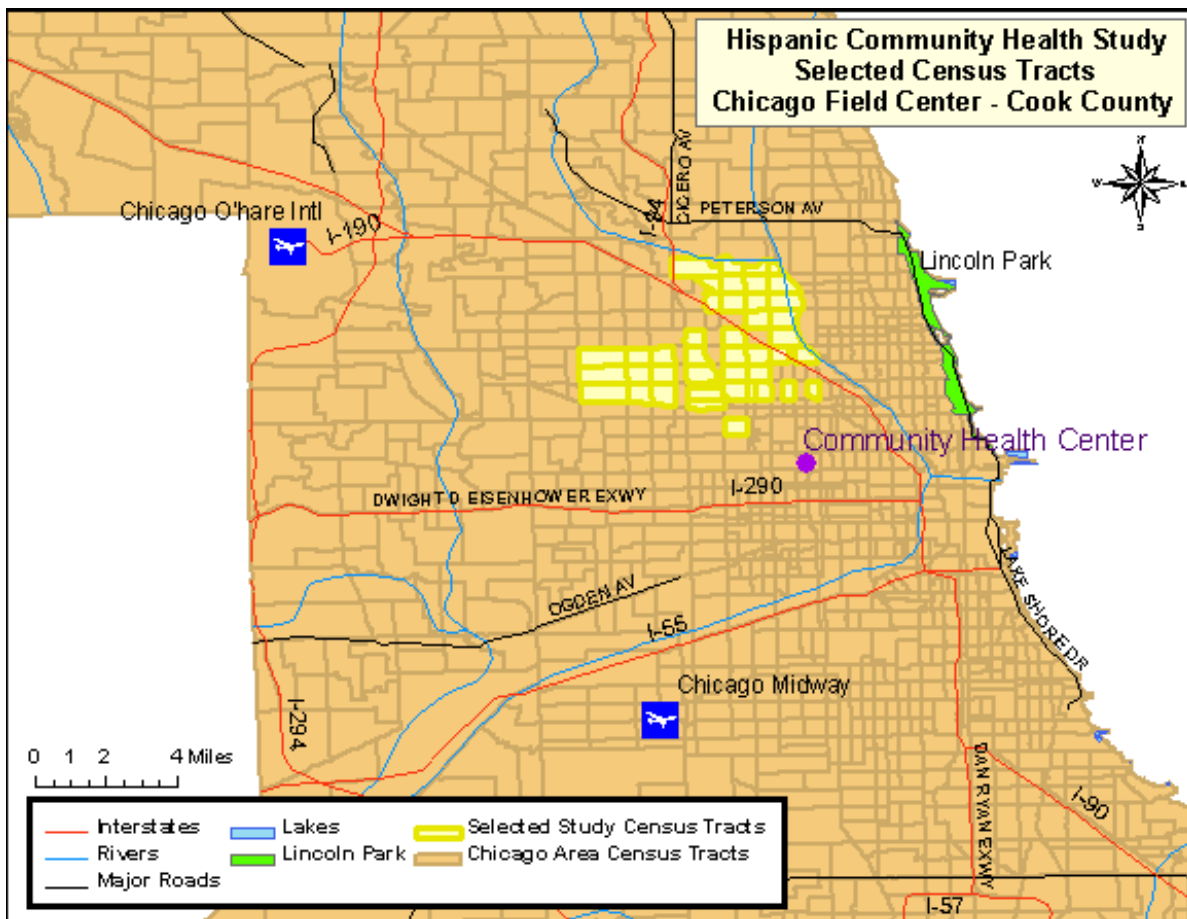
For recruiting, areas of the Bronx that have the highest Hispanic/Latino concentration and that are in closest proximity to the Bronx Field Center location(s) in the South and East Bronx are targeted. Map 1 highlights the specific recruiting areas for the Bronx. Areas highlighted represent the selected census tracts.



Map 1: HCHS/SOL Selected Census Tracts, Bronx

Chicago: over 1.7 million residents of Hispanic/Latino origin

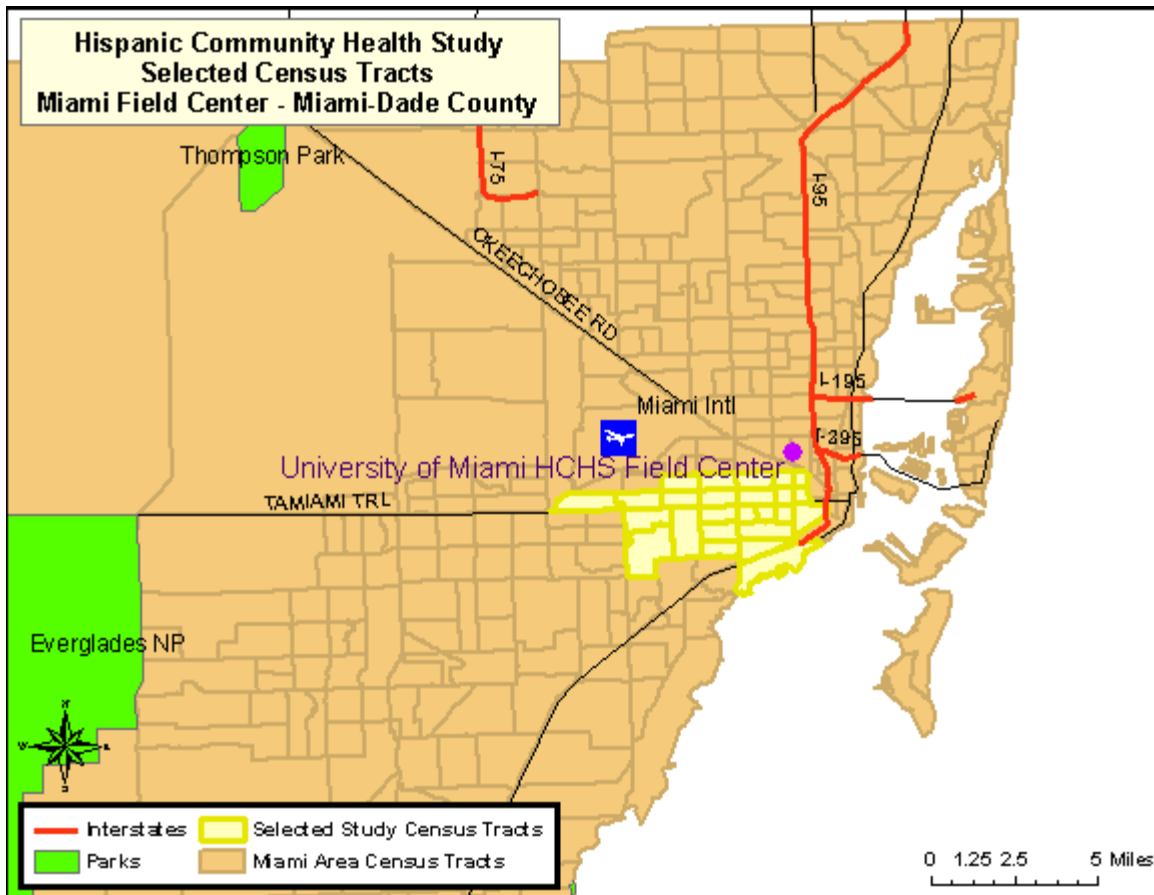
The targeted area for the Chicago site is composed of ethnically diverse neighborhoods with several that have been majority Hispanic/Latino for decades as well as others that were traditionally White/European-immigrant which have experienced Hispanic/Latino in-migration only recently. The highlighted areas in Map 2 represent the selected census tracts in the Cook County, where Chicago is located.



Map 2: HCHS/SOL Selected Census Tracts, Cook County

Miami/Dade County: over 1.3 million residents of Hispanic/Latino origin

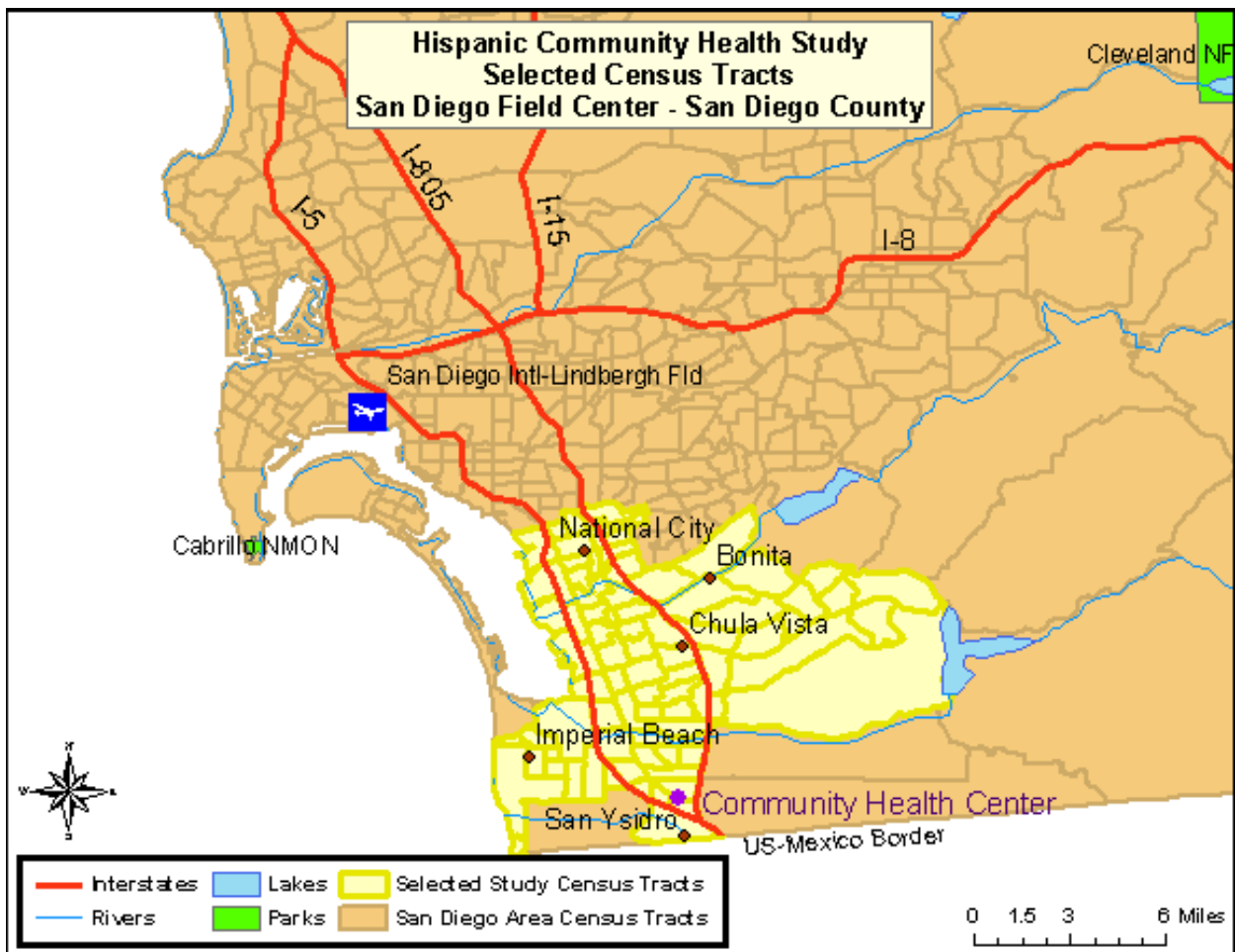
The highlighted areas of Map 3 represent the selected census tracts in Miami-Dade County. This area consists of approximately 20 contiguous census tracts beginning just south of the Miami Field Center and extending further south and west to the city of Coral Gables. Most of the targeted census tracts are located in the city of Miami.



Map 3: HCHS/SOL Selected Census Tracts, Miami-Dade County

San Diego: almost 1 million residents of Hispanic/Latino origin

The combined region of South Suburban and South-Central San Diego County, commonly referred to as the “South Bay”, is the target community. This area includes the communities of San Ysidro, Chula Vista, Imperial Beach, National City, and Bonita. These areas contain large proportions of minority residents, with Hispanics/Latinos representing the largest percentage [US Census Bureau (2005)]. The highlighted areas in Map 4 represent the selected census tracts in the San Diego County.



Map 4: HCHS/SOL Selected Census Tracts, San Diego County

Recruitment was designed to occur in stable, established, communities so that persons can be contacted over time. Each community has a community social infrastructure and organization that enables community support and feedback.

B.1.e. Recruitment Steps

The recruitment plan consists of three basic steps:

- Initial mailings to sampled households describing the study and inviting the household to be screened
- Optional telephone contacts to households with telephone numbers available from the sampling frame
- In person contacts for households without telephone numbers, households unable to be reached through telephone contacts, and households in Field Centers not conducting any telephone screening.

If contact is established with a household through a telephone call, then household screening is conducted via telephone. If a household visit is required to establish contact, then household screening is conducted in person. Three Field Centers plan to use a combination of telephone and in-person screening, while one Field Center (Miami) plans to conduct all screening visits in person. Once eligibility of a household is established, and individual household members who are present are screened for eligibility, clinic visits are scheduled. If not present, individual household members are contacted at a later date by phone or in-person for screening and scheduling of clinic visits.

The field centers have the option in the recruitment procedures to either make the first contact with a prospective household via a phone call, or an in-person visit. The qualifier 'optional' in the 2nd step in the recruitment protocol reflects the fact that at least one site, Miami, plans to use a lead letter followed by home screening visit, thereby skipping the telephone screening step (Step 2). The Miami target area is geographically small and dense, so this plan represents an efficient approach. The other three sites plan to use all three steps at this point, but

may drop the telephone screening if response is not high and move directly to in-home screening after the lead letter.

B.2. Procedures for Information Collection

Data collection for the HCHS/SOL requires questionnaires in each domain of measurement to be available in both English and Spanish versions. Trained, bilingual interviewers will administer the study questionnaires. Questionnaires for which no existing Spanish translations are available are translated by a subcontracting firm, Research Triangle Institute (RTI), with expertise in multilingual instrument development for large-scale surveys. Both new and existing translations are then reviewed by members of the Translation and Validation Subcommittee. This committee includes members from the four field centers, the coordinating center and the project office who are bilingual and represent all four regions of origin for the study (Mexican, Cuban, Puerto Rican, and Central/South American). Scoring sheets are distributed for each translation on which committee members identify problems with specific items. In addition, committee members are asked to rank each item in order of seriousness of translation issues found. The results of the reviews are discussed via teleconference, and a summary of recommended changes to the translation are sent back to RTI for modification.

Prior to study formation, informal discussions were undertaken with staff and community representatives regarding issues related to study design, content, Spanish translation, and cultural issues related to this study. Each community sampled in this study has completely unique Hispanic origin composition, community interaction and resources, cultural influences, Spanish word usage, and cultural history. Thus, these small informal discussions were undertaken separately in each community and constituted a unique set of interactions with communities of

Cuban, Mexican, Puerto Rican, Dominican, and Central/South American influences. Focus groups of more common goals will be undertaken to perfect the questionnaires, and Spanish translation.

Focus groups are planned to occur upon OMB approval and continue for approximately 6 months. The focus groups provide feedback on questionnaire items in order to verify appropriate translation of Spanish idioms and to discuss alternatives should problems be detected in the first several months of use. The total activity would include eleven groups consisting of approximately 7-10 individuals at each field center (4 field centers), lasting approximately 1.5 hours each and discussing approximately 12-15 items in each group. A trained facilitator leads the discussion, and a recording is made for review and scoring after the session has concluded. Results from the focus groups are analyzed and reviewed by committee members, and a summary of any changes needed as a result are forwarded to RTI. Modifications made to the Spanish translation of the questionnaires will be forwarded to OMB.

B.2.b. Cohort Surveillance Component Design

The Study identifies, abstracts, reviews, and validates cardiovascular and pulmonary events (requiring emergency room visit or hospitalization, or based on death information) which occur in the interim between the baseline exam and each subsequent annual follow-up telephone call. Cardiovascular events include myocardial infarction, sudden cardiac death, stroke and heart failure. Pulmonary events include chronic obstructive lung disease and asthma. In more detail, we do the following:

- A. Identify events from the annual follow-up telephone call which provide information that a hospitalization or ER visit took place and the reason for the visit.

- B. Abstract information from these records and enter into the study database.
- C. Validate the diagnosis by review of the abstracted information either by computer or a review committee.
- D. Identify deaths from information obtained at the annual follow-up telephone call and from a review of the vital statistics lists and obituaries from the state in which the community is located. The Coordinating Center (or Field Center if required for confidentiality) is responsible for conducting a match to the National Death Index periodically.
- E. Tabulate cause of death by obtaining, abstracting, and reviewing all relevant information from death certificates. This information will be confirmed by the next-of-kin, coroner, participant's primary physician, nursing home and hospital records.
- F. Review the abstracted information and validate the diagnoses using trained and certified clinicians designated from each Field Center (a morbidity and mortality classification committee).
- G. Ascertain, review, and validate events.

B.3. Methods to Maximize Response Rates and Deal With Non-response

A working Community Relations committee will assist in the development and implementation of a strategic plan to engage in community outreach activities to recruit eligible participants to the study in the targeted geographical areas. These activities can include making brief presentations about the project to community block clubs, local churches (e.g., Pentecostal and *Mita* groups in the Puerto Rican neighborhoods), community-based organizations, consulate offices that represent diverse Latin-American countries, among other strategies.

To maximize response, careful attention has been paid in the planning phase to address the wide range of literacy levels, the wide range of proficiency in English, idiosyncrasies in Spanish according to country of origin, and a lack of familiarity with research. Clear and culturally accessible explanations concerning why the study is being conducted and how the data

will be collected and used will be provided.

Educational level and literacy were factors seriously considered during the development of all the instruments to be used in the study. **It is important to emphasize that all of the questionnaires will be administered verbally by trained interviewers in either English or Spanish.** Because of the wide range of literacy levels are expected among the participants, they will not be asked to read or answer any questionnaires on their own. The interviewer will be able to repeat questions, and in the cases that merit it, participants will receive a card with the scales or alternative answers printed on them, to facilitate their understanding and get more accurate responses.

With permission of the participant, the interviews will be monitored for quality control purposes. Modifications will be made to questionnaires as needed based on experience with the interviews and these quality control checks. Any modifications to the questionnaires will be forwarded to OMB.

Most of the instruments to be used in the study have been used or adapted from other epidemiological studies and, therefore, have been previously validated in their current version. Therefore, for comparability, the language needs to remain consistent. One questionnaire is under copyright. Some of these instruments had been translated and validated in Spanish. For others, a translation was necessary. For this purpose, the Coordinating Center established a contract with an outside company to perform the translations.

The Translation and Validation committee reviewed all the instruments and evaluated the reading level, the quality of the translations (grammatical quality and use of terms that are understood by Hispanics/Latinos of a diversity of origins), and the cultural relevance and appropriateness of the questions. This process of evaluation was not limited to existent versions

in Spanish or translations done for the study. The English versions were evaluated as well. During this process, the committee identified some phrases or words that could have different interpretations or that needed some modification of their reading level. Finally, an outside Spanish scholar and translator, evaluated the final product before its certification.

Due to the occasional medical vocabulary used in the questionnaires, and the variety of idioms in both English and Spanish, the Translation and Validation Committee created a series of definitions for those specific terms. These are the **Question By Question instructions** or “**QxQs.**” If a participant does not understand the meaning of a term, the interviewer will be able to download a menu with the definitions or alternative term (for example, idioms dependent on birth place or community). In consultation with our medical investigators, medical terms need to remain in the questionnaires with appropriate explanations to the interviewers and participants.

Examples of QXQs are attached <Examples of QxQ.doc>

Considerable effort is expended to ensure adequate participation rates among sample members, once selected and identified as eligible. The recruitment protocol consists of advance mailings describing the study and its objectives, followed by telephone contacts. If possible, household screening and selection of household members is conducted via the telephone. For those not responding to the mailing or telephone contacts, in-person screening visits are conducted.

If informational materials have been mailed to the study participant prior to the call, the interviewee is first reminded of the letter and brochure and the staff person reviews this information and answers questions about the study and its procedures, as required. Verification

of eligibility for all study procedures is part of the visit scheduling procedures. Following an explanation of the HCHS/SOL study and the procedures involved, the interviewer requests an opportunity to verify the individual's eligibility for all procedures. The conditions reviewed during this interview (and listed on the form) include pregnancy, the participant's use of a pacemaker or other implanted electronic device, and a screening for conditions that would preclude the periodontal measurements of the dental examination. Study participants who are pregnant are asked to schedule an examination visit at three months after delivery, and to provide a date by which the HCHS/SOL can re-contact them for this purpose.

During this interview staff also inquires about special needs, such as any medical conditions that would affect the examination or the appointment time, difficulties in getting on or off an examination table or dental chair, or impediments in hearing or reading. Arrangements for a safe and comfortable examination visit are made, consulting with the Clinic Manager as appropriate.

Throughout the sampling and recruitment period the Coordinating Center will prepare reports to monitor enrollment by center, gender, country of origin, and age. Recruitment and examination status reports will plot expected vs. actual recruitment rates for each field center, including cumulative and short-term performance. These reports will be available on the study web site to enable the Steering Committee and NHLBI staff, as well as field center personnel, to closely monitor overall recruitment.

The Study will hold periodic conference calls of the recruitment supervisors of each field center with the Recruitment Committee. Initially these calls will be held every other week, until a field center's recruitment tracks at or above the recruitment goal. The purpose of the calls will be to review successful strategies and to facilitate the sharing of materials. A field center with

sub-optimal yield will be asked to develop supplemental or alternative techniques to improve recruitment rates. If a field center were to encounter persistent recruitment difficulties, a monitoring visit will be made.

B.4. Test of Procedures or Methods to be Undertaken

There will be no new procedures or methods of data collection undertaken during the HCSH/SOL. The procedures and methods of data collection are proven and have ~~all~~ been used in previous collections to minimize burden and improve utility. The attachment <Protocol.source.doc> provides the source of procedures/protocols used in this study. The attachment <x.walk.doc> provides the source studies for the questionnaires.

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing

Data

The following individuals were consulted on statistical aspects:

William Kalsbeek, Ph.D. Phone: (919) 962-3249
Director, Survey Research Unit
University of North Carolina, Chapel Hill

Lloyd Chambless, PhD Phone: (919) 962-3264
Collaborative Studies Coordinating Center – University of North Carolina

Lisa LaVange, Ph.D. Phone: (919) 966-8333
Collaborative Studies Coordinating Center,
University of North Carolina, Chapel Hill

The following individuals are responsible for data collection:

Lloyd Chambless, PhD Phone: (919) 962-3264
Collaborative Studies Coordinating Center
University of North Carolina, Chapel Hill

Martha Daviglus, MD, PhD Phone: (312) 908-7967
Chicago Field Center: Northwestern University

Neil Schneiderman, PhD Phone: 305-284-5467
Miami Field Center: University of Miami

Greg Talavera, MD, MPH Phone: 619 594-4086
San Diego Field Center: San Diego State University

Sylvia Wassertheil-Smoller, PhD, FACE, FAHA Phone: (718) 430-2358
Bronx Field Center: Albert Einstein College of Medicine

The following individuals are responsible for data analysis:

Lloyd Chambless, Ph.D. Phone: (919) 962-3264
Collaborative Studies Coordinating Center
Department of Biostatistics
University of North Carolina at Chapel Hill

William Kalsbeek, Ph.D. (919) 962-3249
Director, Survey Research Unit
University of North Carolina, Chapel Hill

Gerardo Heiss, M.D., Ph.D. Phone: (919) 962-3253
Collaborative Studies Coordinating Center
University of North Carolina, Chapel Hill

Lisa LaVange, Ph.D. Phone: (919) 966-8333
Collaborative Studies Coordinating Center,
University of North Carolina, Chapel Hill