

1The Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

Supporting Statement Part B

November 23, 2011

OMB# 0925-0584

Contact:

Larissa Avilés-Santa, M.D.

301-435-0450; avilessantal@nhlbi.nih.gov

6701 Rockledge Drive MSC 7936

Bethesda, MD 20892

FAX: 301-480-1455

Table of Contents

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS	3
B.1. RESPONDENT UNIVERSE AND SAMPLING METHODS	3
B.1.A. DESIGN SUMMARY	3
B.1.B. STATISTICAL CONSIDERATIONS AND POWER	4
B.1.C. SAMPLE SELECTION	7
B.1.D. SAMPLE SIZE REQUIREMENTS	10
B.2. PROCEDURES FOR INFORMATION COLLECTION	14
B.2.A. COHORT SURVEILLANCE COMPONENT DESIGN	15
B.3. METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NON-RESPONSE	15
B.4. TESTS OF PROCEDURES OR METHODS TO BE UNDERTAKEN	18
B.5. INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS AND INDIVIDUALS COLLECTING AND/OR ANALYZING DATA	18

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. Recruitment has been completed as of June 30, 2011. To date, the annual follow-up rate is 85% for the first year and 83% for the second on those portions of the cohort that are past the window of contact. Annual follow-up (AFU) of the full cohort is ongoing and the rate is expected to increase as AFU continues. Recruitment and sampling methods are described in the following publications:

Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglius ML, Giachello AL, Schneiderman N, Raj L, Talavera G, Allison M, Lavange L, Chambless LE, Heiss G. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010 Aug; 20(8):629-41.

<http://www.sciencedirect.com/science/article/pii/S1047279710000724>

Lavange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, Liu K, Giachello A, Lee DJ, Ryan J, Criqui MH, Elder JP. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010 Aug; 20(8):642-9.

<http://www.sciencedirect.com/science/article/pii/S1047279710001171>

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 sample size at the baseline examination by country of origin and community. This table reflects actual recruitment.

<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>	<u>Mixed/ Other</u>	
Bronx	406	45	208	3210	214	4083
Chicago	792	25	2410	799	101	4127
Miami	1504	2272	38	146	114	4074
San Dieg	102	9	3817	41	94	4063
Total	2804	2351	6473	4196	523	16347

The above table represents the numbers of participants who completed the full baseline examination or an abbreviated core examination. The previously approved number of burden hours was not exceeded.

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.

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 complete. It was accomplished through a two-stage area probability sample implemented for each site. At the first stage, a stratified sample of Census block groups was 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 was 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 were selected at a higher rate than those in the lower concentration stratum. This over-sampling was carried out to maximize efficiencies in the field by increasing the probability that a selected household was a Hispanic/Latino household. In addition to these four strata, block groups in the Coop City area were 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 were defined to ensure selection of adequate numbers of households in the respective areas.

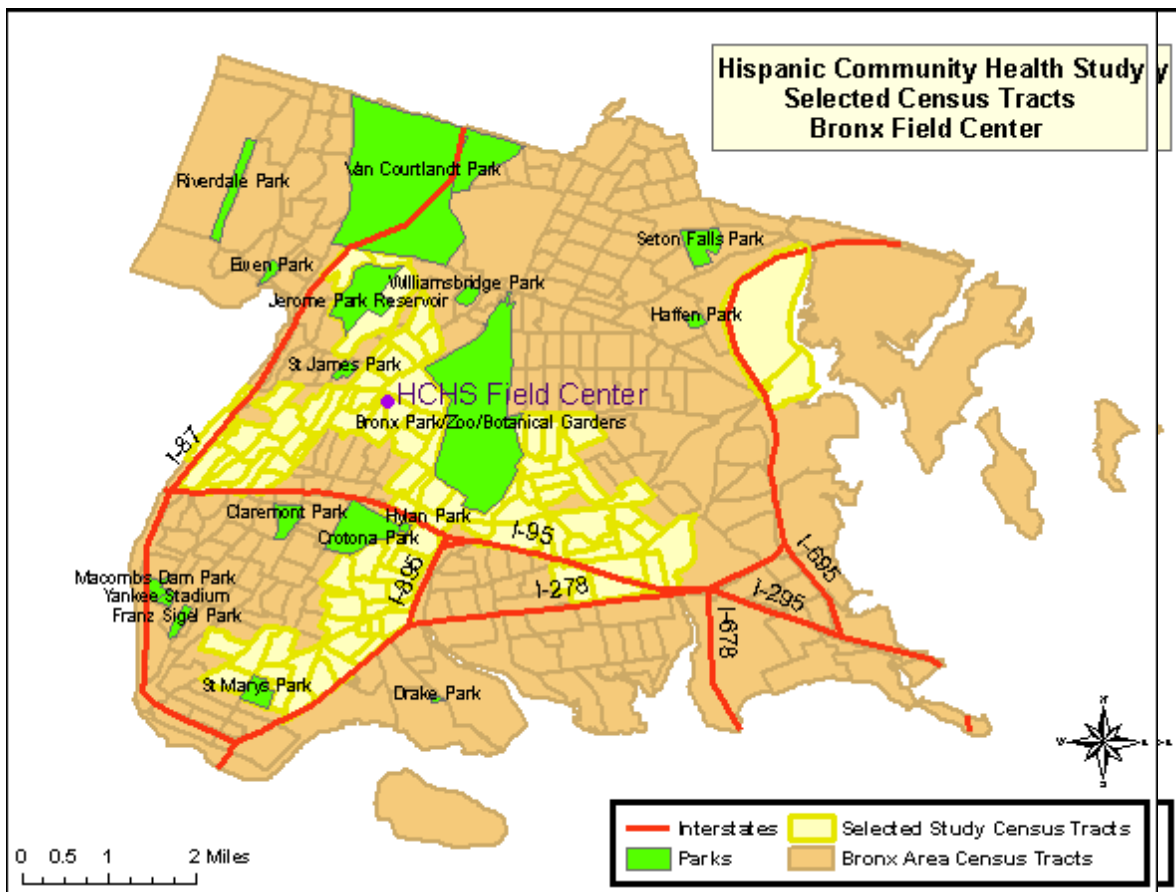
At the second stage, households in the sampled block groups were selected from a dual frame constructed from non-over-lapping lists of postal addresses and Hispanic/Latino surnames. Addresses were selected from the surname list at a higher rate than from the postal list, to further maximize efficiency of field operations by increasing the probability that a selected household was a Hispanic/Latino household. Selected households were 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 was selected, all members of the household were invited to participate. This household selection algorithm was 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 took place over a three-year period. The sample of households in each target area was randomly allocated to each of the three years of recruitment. Within each recruitment year, we fielded 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, were selected from each of four separate communities:

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

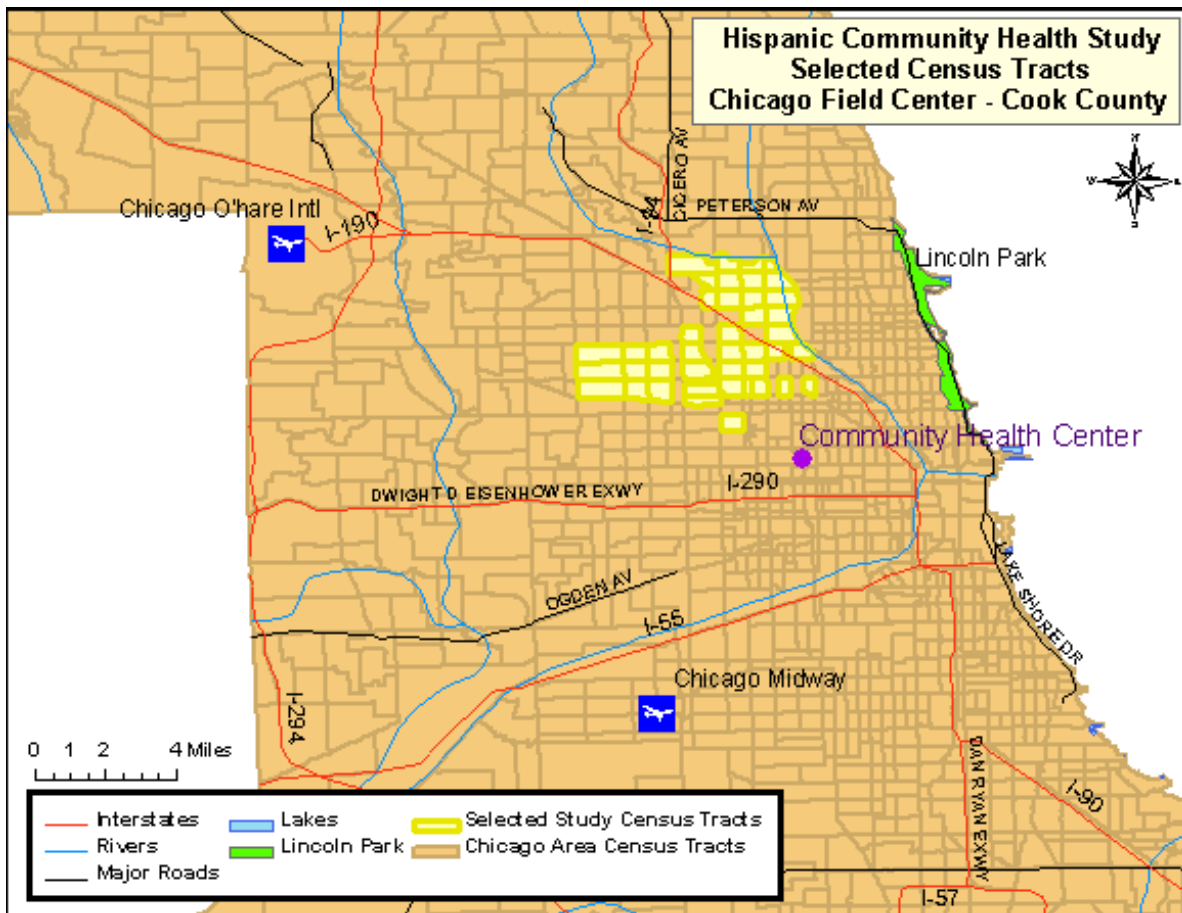
For recruiting, areas of the Bronx that have the highest Hispanic/Latino concentration and that are in closest proximity to the Bronx Field Center location(s) in the South and East Bronx are targeted. Map 1 highlights the specific recruiting areas for the Bronx. Areas highlighted represent the selected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



Map 1: HCHS/SOL Selected Census Tracts, Bronx

Chicago: over 1.7 million residents of Hispanic/Latino origin

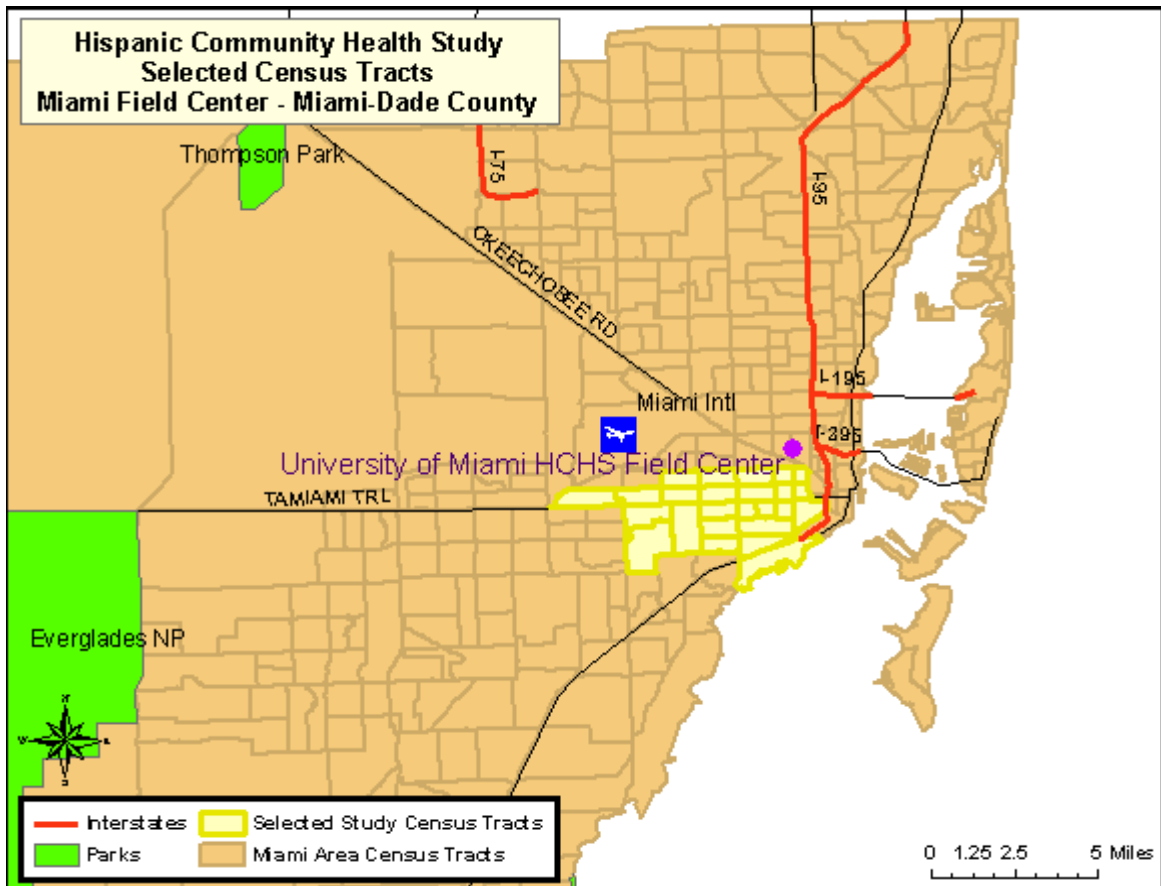
The targeted area for the Chicago site is composed of ethnically diverse neighborhoods with several that have been majority Hispanic/Latino for decades as well as others that were traditionally White/European-immigrant which have experienced Hispanic/Latino in-migration only recently. The highlighted 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

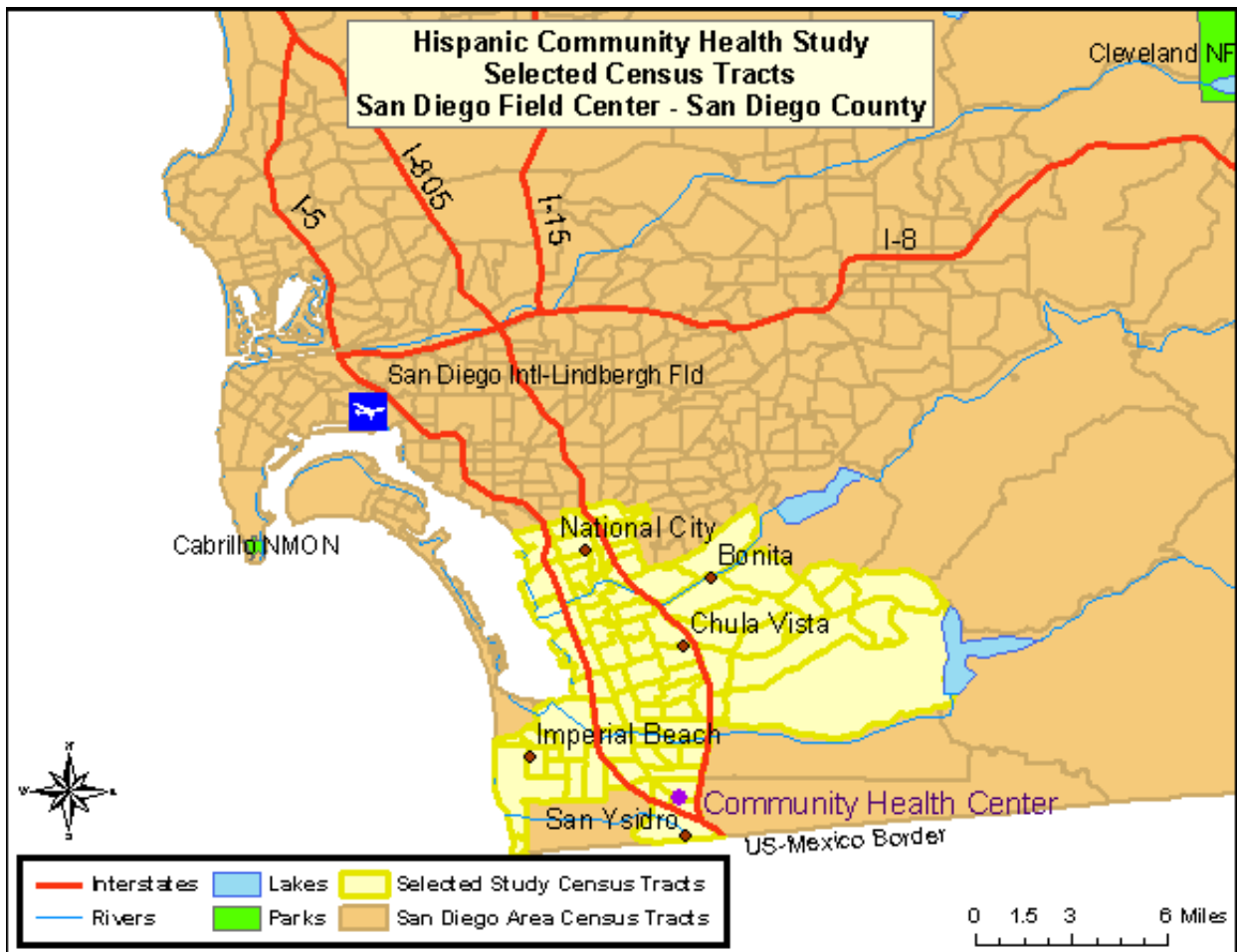
The highlighted areas of Map 3 represent the selected census tracts in Miami-Dade County. This area consists of approximately 20 contiguous census tracts beginning just south of the Miami Field Center and extending further south and west to the city of Coral Gables. Most of the targeted census tracts 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.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 administered the study questionnaires. Questionnaires for which no existing Spanish translations were available were translated by a subcontracting firm, Research Triangle Institute (RTI), with expertise in multilingual instrument development for large-scale surveys. Both new and existing translations were then reviewed by members of the Translation and Validation Subcommittee. This committee included 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 were distributed for each translation on which committee members identify problems with specific items. In addition, committee members were asked to rank each item in order of seriousness of translation issues found. The results of the reviews were 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.

B.2.a. 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

The study has Participant Retention and Community Relations committees whose goals

are to maximize retention of study participants throughout the follow-up period. To best retain HCHS/SOL participants, the study has at least one contact with participants every quarter (i.e. every 3 months). Contacts include post-visit thank you cards or calls, quarterly newsletters tailored to the community of each Field Center, a birthday or greeting card, and a holiday or end-of-year card. These are informational materials and do not impose an information collection burden on the participant. The annual follow-up phone call is accounted for in the burden estimate. Contacts are initiated by each field center and are culturally appropriate, are in the participant's preferred language (English or Spanish). Radio/TV public service announcements in the local the communities will serve as reminders about the study for participants and the community at large.

For annual follow-up, all living participants who met the minimal standards for the baseline examination have been contacted annually unless they have specifically requested no further contact. This includes participants who have moved away from the community in which they were recruited. Study participants are contacted as closely as possible to their baseline examination anniversary date. Repeated phone attempts are made at different times of the day, and home visits are scheduled if needed. Annual follow-up contacts began in March, 2009 since that is one calendar year since the start of the baseline examinations in 2008.

If the participant is not available or unable to respond, an alternate respondent designated by the participant is contacted.

To maximize response, careful attention was paid in the planning phase to address the wide range of literacy levels, the wide range of proficiency in English, Spanish idioms and regionalisms, and a lack of

familiarity with research.

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 were administered verbally by trained interviewers in either English or Spanish.** Because of the wide range of literacy levels were expected among the participants, they were not asked to read or answer any questionnaires on their own. The interviewer was able to repeat questions, and in the cases that merit it, participants received 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 were monitored for quality control purposes. Most of the instruments used in the study were used or adapted from other epidemiological studies and, therefore, have been previously validated in their current version. Therefore, for comparability, the language needed to remain consistent. 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 translations of the instruments and evaluated the reading level, the quality of the translations (grammatical quality and use of terms that are understood by Hispanics/Latinos from a diversity of origins), and the cultural relevance and appropriateness of the questions. The English versions were evaluated as well. 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 did not understand the meaning of a term, the interviewer was 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 needed to remain in the questionnaires with appropriate explanations to the interviewers and participants.

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

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

There will be no new procedures or methods of data collection undertaken during the HCHS/SOL.

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, Ph.D. Phone: (919) 962-3264
Collaborative Studies Coordinating Center – University of North Carolina

Lisa LaVange, Ph.D.
Collaborative Studies Coordinating Center - University of North Carolina, Chapel
(recently moved to a position outside UNC)

Jianwen Cai, Ph.D. Phone: (919) 966-7788
Collaborative Studies Coordinating Center,
University of North Carolina, Chapel Hill

The following individuals are responsible for data collection:

Jianwen Cai, Ph.D. Phone: (919) 966-7788
Collaborative Studies Coordinating Center
University of North Carolina, Chapel Hill

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

Neil Schneiderman, Ph.D. 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

Robert Kaplan, Ph.D. Phone: (718) 430-4076
Bronx Field Center: Albert Einstein College of Medicine

The following individuals are responsible for data analysis:

Jianwen Cai, Ph.D. Phone: (919) 966-7788
Collaborative Studies Coordinating Center,
University of North Carolina, Chapel Hill

William Kalsbeek, Ph.D. Phone: (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