3. SCREENING

Eligibility

Men and women must meet the following criteria to participate in the Baseline Interview.

- Primary residence in a designated community within the city of Boston, Massachusetts.
- Able to speak and read English or Spanish
- Age 30-79 YEARS at time of first contact.
    - For preselected contact, first contact is the Household screener
    - For other household members, first contact is the Individual Screener
- Cognitively able to provide verbal consent.
- Self-identified as Hispanic regardless of racial self identification
- Self-identified race in the following areas
    - Respondent considers him/herself exclusively African-American
    - Respodent considers him/herself African-American, mullti-racial
    - Respondent considers him/herself exclusively White

Sampling Design

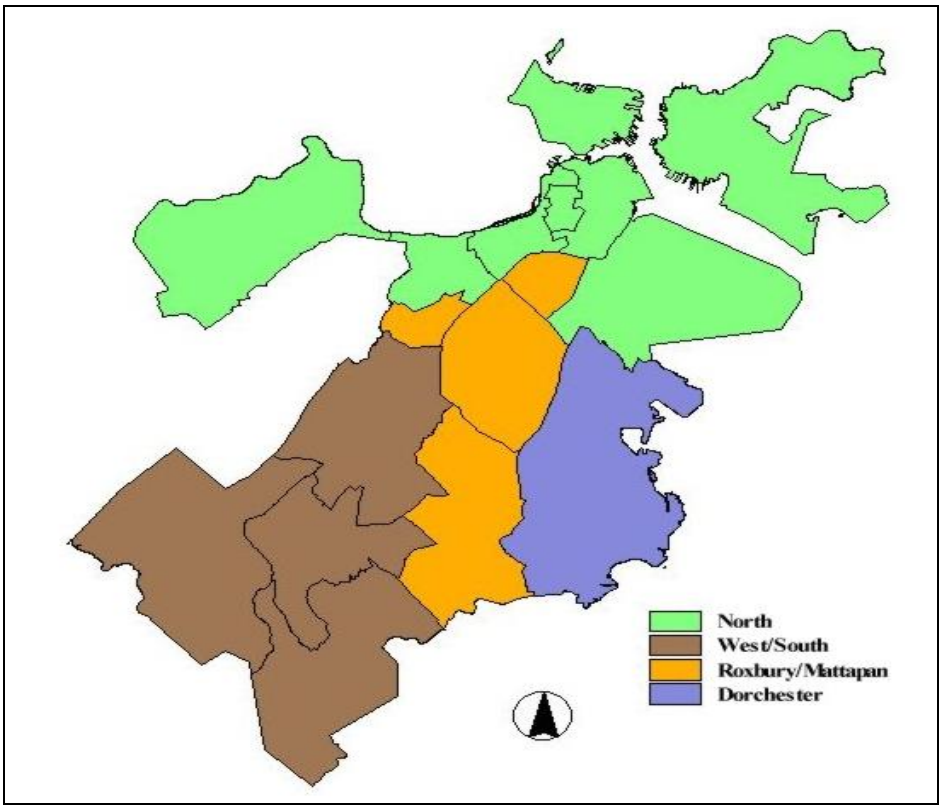
Overview. Our objective is to draw representative samples of men and women (aged between 30 and 79 years) in three racial and ethnic groups from all areas of Boston. A stratified cluster sample is proposed to achieve this distribution, where census blocks are the cluster unit. Some sampling will be done within census blocks (especially for younger non-Hispanic individuals), as required, to obtain the desired age and ethnic balance. Census blocks in Boston are classified into 12 strata, with sampling fractions varying by stratum.

Stratification.

(a) Neighborhood groups. There are 16 major residential planning districts (or neighborhoods) in Boston (Boston Redevelopment Authority, 2001). These neighborhoods were grouped into the following four major geographic areas (see Figure 2) so as to produce more racially/ethnically homogeneous sub-areas.

1. North (Allston/Brighton, Back Bay/Beacon, Central, Charlestown, East Boston, Fenway, South Boston);
2. West/South (Hyde Park, Jamaica Plain, Roslindale, West Roxbury);
3. Roxbury/Mattapan (Mattapan, Roxbury, South End); and
4. Dorchester (North and South Dorchester).

Boston Neighborhoods, Grouped by Geography and Race/Ethnic Diversity



(b) Ethnic/Racial Densities. Within each of the 4 geographic areas, Census blocks were classified as:

1. Low density African American and Hispanic, or areas in which less than 25% of the residents are African American and less than 30% are Hispanic
2. High density African American, or areas in which 25% or more of the residents are African American
3. High density Hispanic, or areas in which 30% or more of the residents are Hispanic.

The classification scheme yields 12 strata (4 geographic areas x 3 density categories).

Table 1 shows the distribution of the Boston population (according to the 2000 census) by gender, race/ ethnicity, and age.

Table 1: Distribution of Boston Population by Gender, Race/Ethnicity, and Age Decade

Gender	Race/Ethnicity	Age					Total
		30-39	40-49	50-59	60-69	70-79	
Male	Hispanic	7548	4687	2441	1124	485	16285
Male	Non-Hispanic African American	10794	8978	6069	3516	1959	31316
Male	Non-Hispanic Caucasian	28177	18618	13773	9122	7745	77435
Female	Hispanic	7951	5080	2912	1571	844	18358
Female	Non-Hispanic African American	13475	11177	8229	5047	3297	41225
Female	Non-Hispanic Caucasian	24701	17367	14221	10505	11465	78259

We compared various sampling fractions for the 12 strata to determine how best to obtain the sample outlined in Table 1 given the distribution outlined in Table 2. This comparison of various sampling fractions suggested that Caucasian representation could be achieved without sampling any of the low-density blocks. However, we rejected this approach, as the sample would not be representative of the city of Boston as a whole. We calculated an "ideal" distribution of sample blocks, if unlimited resources were available, for a self-weighted sample. This yielded a sampling fraction of 45 percent for each stratum. The design factor, 1 + L, as defined by Kish, 1965, gives the increase in inefficiency (relative increase in variance) by not selecting a self-weighted or simple random sample (SRS).

$$\sigma^2_{\text{BACH design}} / \sigma^2_{\text{SRS}} = 1 + L = \frac{\sum_{h=1}^{12} P_h \cdot W_h^2}{\left(\sum_{h=1}^{12} P_h W_h \right)^2}$$

where: P_h = the proportion of the sample of interest selected from stratum h;
W_h = the estimation weight value for subjects selected in stratum h.

Finding adequate Hispanics, especially for older males, is the constraining factor for any design. We decided to choose stratum sampling fractions that would allow the average design factor to be no more than 1.5 and the design factor within each stratum to be no more than 2. There are four age, two gender, and three ethnic categories for a total of 24 design groups (see Table 1). Choosing as sampling fractions, across all four regions:

- .10 for the low-density areas;
- .10 for the African American areas; and
- .70 for the Hispanic areas;

meets our criteria.

Table 2 reports the total number of blocks that will be sampled from each of 12 strata given our sampling fractions. For example, with these sampling fractions, we will sample about 118 (10%) of the 1184 available blocks in the predominantly Caucasian area of North Boston. Since we intend to break up the sample into 6 equal randomly selected batches (see Section 4), this means we will sample 19 blocks in this area per batch. By contrast, we will sample 131 blocks (70%) of the 187 available blocks in the areas of North Boston where 30% or more of the residents are Hispanic.

The sampling will occur by assigning each of the 4,266 blocks in Boston a uniform random number between 0 and 1. We will then group the blocks according to stratum and, within each stratum, order the blocks sequentially from lowest to highest random number assigned. We will then select 10,10, or 70% of the blocks (depending on the stratum), beginning with the smallest random number.

Table 3 shows the total number of *potential* survey respondents available in the 788 sampled blocks. We recognize, however, that not all residents contacted will be willing to participate in the study. If we expect that 50% of known, eligible subjects contacted will participate, and assume that 65-70 percent of subjects contacted will be eligible then we would like our numbers in Table 4 to be three times those in Table 1.

Table 2. Total and Sampled Number of Blocks Per Strata

Strata							
Number	**Atea**	**Race/Ethnicity**	**Total Number of Blocks**	**Multiplied By**	**Total # of Sampled Blocks**	**Divided By**	**# of Sampled Blocks per Batch**
1	North	Low-density	1184	.1	118	6	19
2	North	High African American	23	.1	2	6	1
3	North	High Hispanic	187	.7	131	6	22
4	West/ South	Low-density	942	.1	94	6	16
5	West/ South	High African American	207	.1	21	6	4
6	West/ South	High Hispanic	169	.7	118	6	19
7	Roxbury/ Mattapan	Low-density	123	.1	12	6	2
8	Roxbury/ Mattapan	High African American	592	.1	59	6	10
9	Roxbury/ Mattapan	High Hispanic	171	.7	120	6	20
10	Dorchester	Low-density	337	.1	34	6	6
11	Dorchester	High African American	255	.1	26	6	4
12	Dorchester	High Hispanic	76	.7	53	6	9
TOTAL			4266	.	1788		132

Table 3: Potential Sample Sizes* in 24 Demographic Groups					
		Age			
Gender	**Race/Ethnicity**	**30-39**	**40-49**	**50-59**	**60-79**
Male	Hispanic	3129	1910	1016	625
Male	Non-Hispanic African American	2094	1680	1122	999
Male	Non-Hispanic Caucasian	3893	2771	1995	2447
Female	Hispanic	3355	2146	1216	930
Female	Non-Hispanic African American	2827	2200	1639	1647
Female	Non-Hispanic Caucasian	3509	2625	2096	3303

* Assumes no sampling within census blocks

Development of the Sampling Frame. We have considered two approaches to developing a sampling frame from which a random sample of community-dwelling subjects can be selected. They are as follows:

1. Door-to door enumeration of all households within the sampled Census blocks; and
2. Selection of individuals from the Massachusetts Resident Lists (Boston) for the sampled Census blocks.

The project originally proposed to use the first method (door-to-door enumeration) but this is now not feasible given the substantial budget reduction recommended by the review group. We now propose to use the second method (Resident Lists), which has its own advantages as follows:

- The Massachusetts Residents Lists are relatively (about 90%) complete (Bohlke et al., 1999) and up-to-date. The Election Commission of each city or town updates their Resident List annually. They do this by sending forms to each address, requesting the names, genders, dates of birth, occupation, and voter status of each resident aged 17 and older as of January 1, that year. Residents are required by law to return these completed forms; if they do not, the Election Commission typically sends a data collector to the residence;
- NERI has used the Massachusetts Resident Lists successfully in previous studies. Such studies include the Massachusetts Male Aging Study (MMAS) and the Massachusetts Women's Health Study (MWHS);
- The Boston Resident List can be geocoded easily and with up to 84% accuracy (Krieger et al., 2001).
- The Boston Resident List enables us to identify a specific person to contact with an introductory letter explaining the study and providing documentation *before* a NERI data collector calls or visits to screen them for eligibility. We expect that this

introductory letter will encourage more cooperation than we would get through an enumeration.

- About half of the people identified through the Boston Resident List will have listed telephone numbers, which means data collectors can call to schedule visits, saving time and travel costs.

Methodological Sub-Study. Despite the advantages of using the Boston Resident List for participant recruitment (Method 2), we plan to conduct, as a precaution, a methodological sub-study comparing it to a small-scale enumeration (Method 1). Specifically, we will sample ten blocks from the four major geographic areas in Boston, including two predominantly African American, seven predominantly Hispanic, and one predominantly Caucasian block. These blocks will be part of the first Batch sub-sample. Both objectives will be addressed by halving each of the ten blocks and randomly assigning each recruitment method to half.

The objectives of the methodological sub-study will be to:

1. Assess the level of agreement between the two identification and recruitment methods; and
2. Determine and compare the costs for identifying and making appointments for baseline interviews with eligible men and women using Method 1 compared to Method 2.

Objective 1, the assessment of agreement (coverage) between the two methods, will be addressed first. A list of the names and addresses of eligible men and women found in designated half of the sampled blocks will be compiled using Method 1 (the enumeration). Independently of this enumeration, we will print the list of age-eligible men and women in the same blocks from our electronic Boston Resident List file. We will then assess the level of agreement between the two identification methods, for the half of block for which we have enumeration and list information, by taking note of names found by both methods, names found only by Method 1, and names found only by Method 2. To illustrate, if there are approximately 45 people in our target age range per block, this analysis should be based on a total of about 225 names. Method 2 will be considered to be acceptable if it finds at least 85% of all people identified by either method.

Objective 2 Method 1 (enumeration) will involve travel, time to identify household units, time to screen and enlist available individuals, and a follow-up telephone call to screen and enlist people identified through the enumeration process but not available. When Method 2 (Resident List) can make use of the telephone to contact people (i.e., when the sampled person has a listed number), then an appointment for the full interview will be scheduled in this way. When Method 2 cannot make use of the telephone (i.e., the sampled person does not have a listed number), travel costs will be involved for a field screening visit. Thus both methods will utilize a combination of telephone and field screening to determine eligibility and schedule appointments. The relative costs of these two methods will depend primarily on the ratio of telephone to field screening.

The sampling process for both the enumeration and Resident List methods of developing the sampling frame will be straightforward. Each individual in the Boston Resident List will be assigned a uniform random number of between 0 and 1 before beginning the sampling. For the

enumeration portion of this study, we will provide lists of random numbers and the enumerator can assign the random numbers sequentially to those individuals enumerated who are between 30 and 79 years.

Sampling within Census Blocks

Looking at Table 3 above, some cells contain greater numbers of potential respondents than are needed for this study. To reduce the sample in these cells, cut points in the random number can be determined for each gender/race/ethnicity/age cell so that a more appropriate number of people are contacted. If we know an individual's age from the Resident List, we can eliminate the initial contact if their random number is greater than the maximum of the cut-off points for their gender and age. Only those whose random number is below the cut-point for their gender/race/ethnicity/age cell will continue with the interview. Table 4 gives the cut points for the random numbers, for a response rate of 1/3. As an example, in the study a block is chosen for sampling. Within a block there are 20 individuals who are eligible for sampling. The first individual is a 35-year-old male with a random number of .54. Given that this random number is greater than .3582, this individual is not contacted. The second individual is a 37 year old female with a random number of .24. This random number is less than .2653 so the individual is screened. If the individual is non-Hispanic African American then they are eligible, if other ethnic group, the individual is not eligible.

Table 4: Cutoff points for random number (upper bound on usable random number in each design group)

		Age			
Gender	**Race/Ethnicity**	**30-39**	**40-49**	**50-59**	**60-79**
Male	Hispanic	0.2397	0.3927	0.7382	1
Male	Non-Hispanic African American	0.3582	0.4464	0.6684	0.7510
Male	Non-Hispanic Caucasian	0.1927	0.2701	0.3759	0.3065
Female	Hispanic	0.2235	0.3495	0.6168	0.8066
Female	Non-Hispanic African American	0.2653	0.3409	0.4576	0.4554
Female	Non-Hispanic Caucasian	0.2137	0.2857	0.3578	0.2271

The cut-offs can be adjusted if necessary as the study proceeds (see Section 4, on "batching").

Sampling Implementation – "Batching"

Following successful experiences on other large-scale epidemiologic projects, we plan to implement the sampling design in "batches", or successive random subsamples (Feldman, McKinlay and Niknian, 1996) of 379 Census blocks across 12 strata as shown in Table 2. Each

batch will yield final samples of approximately 500 men and 500 women. Attempts will be made to contact all eligible subjects within the Batch in the first quarter (i.e., first 3 months) that that batch is active. It is planned that 90-95% of all final dispositions will be made during this quarter. During the second quarter of each batch, most of the remaining difficult dispositions will be pursued. Quality control and data entry will be ongoing in both quarters. Thus, each batch will be completed in a 6-month period.

The first stage of implementation will be to draw a random sample of blocks from each stratum. Within each stratum, blocks will be placed in a random order and selected in sequence. Within each selected block, we will contact all men 70-79 years old and take samples of younger men to produce target numbers of respondents in each age group. In order to obtain approximately equal numbers of subjects in each racial/ethnicity category, the fraction of all blocks to be sampled will differ from stratum to stratum. Blocks in minority areas will be oversampled and those in predominantly non-minority areas will be undersampled.

There are distinct advantages to this batching approach:

- It ensures a manageable schedule for the extensive fieldwork (n = 6000);
- It adds a random factor to provide estimates of short-term community, seasonal and individual level fluctuations (e.g. hormones, physical activity levels, and participant availability);
- Each batch will provide valuable experience for subsequent batches (e.g., the experience of the first batch may suggest alterations in the subsequent sampling approach due to logistical challenges);
- It permits us to monitor response rates by age, gender, and race/ethnicity as the study progresses, and revise sampling fractions as needed.
- It reduces field staff burnout by distributing difficult dispositions (refusals and non-contacts) across the entire data collection period; and
- Since each Batch is a random sub-sample, it is easy to add small ancillary studies as promising new research ideas emerge (e.g. ED and subclinical CVD, hormones and osteoporosis in men, endometriosis). Already well-characterized subjects will be available (after obtaining appropriate subject consent).