Supporting Statement B

for

The Healthy Communities Study: How Communities Shape Children's Health (NHLBI)

September 14, 2011

Name: Sonia Arteaga
Address: 6701 Rockledge Drive, Suite 10018, Bethesda, Maryland 20892
Telephone: (301) 435-6677
FAX: (301) 480-5158
Email: arteagass@nhlbi.nih.gov

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ATTACHMENTS

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Respondent Universe and Sampling Methods

B.1.a Design Summary

The Healthy Communities Study (HCS) will collect both current/prospective and retrospective data on 23,322 child participants and their parents (20,358 in the first three years of data collection) in 279 communities across the United States in two waves. The four Wave 1 communities will be selected for convenience in the Baltimore/Washington DC area because the focus of Wave 1 is protocol refinement. In Wave 2, the HCS will use a hybrid design that includes a stratified national probability-based sample (NPBS) of 195 communities combined with a sample of 80 communities selected with certainty that have promising programs and policies aimed at reducing childhood obesity.

B.1.b Respondent Universe – Communities

In the HCS, a community is defined as the catchment area of a local public high school. The HCS will use a hybrid design in Wave 2 that involves a stratified NPBS of 195 communities combined with a sample of 80 communities selected with certainty to represent communities with promising programs and policies aimed at reducing childhood obesity. Selection of the four communities for Wave 1 will be conducted before the Wave 2 communities are selected. Wave 1 communities will be selected for convenience to be in reasonable proximity to Battelle and NHLBI offices in the Baltimore/Washington DC area while allowing for diversity with respect to urbanicity, proportion of minorities, and income. The following subsections provide detail on the selection of Wave 2 communities into the HCS.

National Probability-Based Sample of Communities: For the 195 Wave 2 communities selected in the HCS, we will employ a stratified NPBS. The NPBS ensures that the HCS can be used to provide nationally representative estimates when conducting weighted analyses data. Each stratum in the design will represent a group of census tracts with similar characteristics with respect to geographic region, urbanicity, race/ethnicity, and income. We will also stratify census tracts based on a pre-selection scoring of childhood obesity programs and policies. Methods used to establish this pre-selection scoring are described in Section B.1.b.1.

Each census tract across the United States will be assigned to one stratum. Within each stratum, a single census tract will be selected using a probability-based selection approach. We will then determine which high-school catchment area contains the highest fraction of the population within the selected census tract to select the community into the study.

This method of selecting census tracts and communities within counties will help ensure all communities across the United States have a known probability of being included in the HCS, and when combined with the certainty communities, will yield a nationally representative study. Construction of the strata will ensure representation across a range of geographically, demographically, and culturally diverse communities.

Certainty Communities: Eighty certainty communities are included in the hybrid design to ensure inclusion of communities that have a history of sufficiently promising program and policy initiatives to reduce childhood obesity. The approaches used in selecting these communities are described below.

B.1.b.1 Processes for Gathering Information for Community Selection and Stratification
The project team will develop and populate a comprehensive and hierarchical Geographical Information Systems (GIS) database of states, counties, municipalities, and neighborhoods that have policies and programs targeting childhood obesity. A hierarchical structure is needed because programs and policies will be implemented at differing levels of geographic specificity, most of which will not be consistent with the high-school catchment area definition of a community that is being used in this study. GIS software will be used to associate census tracts, school districts, and high-school catchment areas with broader geographic boundaries and vice versa. The database will be used for the selection of certainty communities, the stratification of counties in the design, and to leverage an existing GIS database (Advancing the Movement) developed with support from the National Collaborative on Childhood Obesity Research (NCCOR).

The GIS database will be used to develop a pre-selection scoring of communities and counties. Variables that will contribute to this pre-selection scoring include the following:

- Annualized per-capita funding received for childhood obesity program implementation (population size defined by Census 2010 Population estimates for the number of children aged 3-15 years)
- Duration of programs and policies identified
- Whether the program or policy was developed and implemented at the community, county or state/regional level
- Number of programs and policies identified
- Number of independent nominations received by members of the HCS Steering Committee

To populate the GIS database, multiple approaches will be employed while using a standardized information-gathering template as shown in Attachment 1. This template will be used to gather pre-selection information on programs and policies that occur at different levels of geographic specificity (state, county, municipality, or neighborhood). The sources of data that we will use to populate the GIS database are listed below:

1. ***HCS Steering Committee Members***: We will include recommendations/nominations of specific communities to be considered as candidates for certainty communities from members of the HCS Steering Committee (or their designees). The person nominating a community will provide information using the template shown in Attachment 1. A variable will be captured in the GIS database to indicate that a particular community was recommended by one or more members of the Steering Committee – one of the factors that is used to stratify communities as described below in Section B.1.b.2.

2. ***Federal Funders***: We will perform a census of the large federally supported programs, such as the National Heart, Lung, and Blood Institute's (NHLBI's) ***We Can!*** (**W**ays to **E**nhance

Children's **A**ctivity & **N**utrition)® program, and the Centers for Disease Control and Prevention's (CDC's) Communities Putting Prevention to Work (CPPW) program, to gather information on program characteristics (i.e., community name, level of geographic specificity funded, level of funding, duration of funding, and general information about the program goals).

3. ***Non-federal Funders***: We will contact members of the Convergence Partnership, a collaboration of large funders of childhood obesity programs, to gather data on programs they fund. The Convergence Partnership includes the California Endowment, Kaiser Permanente, the Kresge Foundation, Nemours, the Kellogg Foundation, and CDC. A maximum of nine non-federal entities will be contacted prior to Office of Management and Budget (OMB) approval. We will ask program funders to complete the information-gathering template for communities that were funded to reduce childhood obesity over the past decade (i.e., community name, level of geographic specificity funded, level of funding, duration of funding and general information about the program goals).

4. ***Online Policy Databases***: We will research national, state and local policies that affect childhood obesity, such as those adopted by the state of Arkansas to promote BMI evaluation and appropriate nutritional and physical activity follow-up among school-aged children. There are multiple documented public databases available, which provide information about such policies, including nutritional and physical activity policies in schools (e.g., the National Cancer Institute's (NCI's) Classification of Laws Associated with School Students database). These policies will also be incorporated into the GIS database, including duration of policy and level of geographic specificity (school district, town/city, county, state).

5. ***State Health Departments***: Following OMB approval, we will survey leaders involved in combating childhood obesity within each state (most likely the head of the Chronic Disease Prevention program within the State Health Department). We will ask these state leaders to complete the information-gathering template in Supporting Statement A Attachment 18.

6. ***General Program/Policy Information***: In data gathering (before and after OMB approval), we will conduct initial outreach via e-mail to set up telephone appointments with funders/sponsors of programs and policies (e.g., YMCA) to gather general information, key documents (via e-mail, mail, links to Websites, etc.), funding levels, and duration of funding within specific geographic areas via databases. We will ask these organizations to provide us with available information regarding their programs and policies, but we will not ask them to complete any sort of questionnaire. Thus, there will not be any participant burden from these individuals.

B.1.b.2 Selection of Communities for Nationally Representative Sample

We plan to select the first 195 communities in Wave 2 using a stratified probability-based approach to make inferences based on nationally representative estimates from the study. Following this approach, we will group census tracts into strata based on race/ethnicity, urbanicity, income, region of the county, and pre-selection score – followed by selection of census tracts from the strata using a probability proportional to size algorithm (where size refers to the number of children aged 3-15 years in each census tract based on Census 2010 data).

After each census tract is selected, our team will termine which high-school catchment area contains the majority of the population from within the selected census tract to identify the specific community for inclusion into the study.

It should be noted that the research team investigated multiple options for drawing the sample, with a strong preference of stratifying the nation based on high-school attendance boundaries. Unfortunately – these boundaries change frequently based on district needs (e.g., efficient assignment of students to different schools within a district) – and there is no current universal data source that captures this information systematically across the entire US. Therefore, it would be difficult to develop an appropriate sampling frame for the study based on high-school catchment area. We considered a two-stage sampling scheme – with sampling at the first stage at the county or school district level, followed by identification of high-school catchment areas within selected counties or district at the second stage of sampling. Our investigation into this method revealed that the level of specificity in our stratification variables (race/ethnicity, urbanicity and income) became diluted within these larger areas (counties and districts), leading to suboptimal stratification.

Therefore, the stratified sampling of census tracts was selected as the best option for drawing an appropriate sample of communities into the study.

The stratification variables used to classify census tracts are detailed further below:

1. Region of the Country

Region	States included	HHS Regions
1 (Northeast)	CT,DE,DC,MA,MD,ME,NH,NJ,NY,PA,RI,VA,VT,WV	I, II, and III
2 (Midwest)	IA,IL,IN,KS,MI,MO,MN,NE,OH,WI	V and VII
3 (Southeast)	AL, FL,GA,KY,MS,NC,SC,TN	IV
4 (Central)	AR,CO,LA,MT,ND,NM,OK,SD,TX,UT,WY	VI and VIII
5 (West)	AK,AZ,CA,HI,ID,NV,OR,WA	IX and X

2. Urban / Suburban / Rural designation

   The methodology we are utilizing for urban/suburban/rural designation is based on the Rural-Urban Commuting Area (RUCA) system which was developed by the U.S. Departments of Agriculture and Health and Human Services and a multi-state/university collaborative in the 1990s. This system uses commuting patterns to distinguish between urban, suburban, large rural, and small rural areas (vs. the more simplistic urban/rural designation provided by the Census). The RUCA system classifies each census tract based on the following characteristics:

   - **Urban Core** areas are contiguous, built-up areas containing 50,000+ people – these correspond to the Census Bureau's Urbanized Areas.
   - **Suburban** areas have high commuting flows to Urban Core areas. These include areas in which 30-49% of the population commutes to Urban Core areas for work.
   - **Large Rural Towns** have a population between 10,000 and 49,999 people and include surrounding rural areas with at least 10% primary commuting flows to these towns, as well as secondary commuting flows of 10% or more to Urban Core areas.

- **Small Towns** and **Isolated Rural Areas** include towns with populations <10,000 people and their surrounding areas, as well as isolated rural zones more than 1 hour (by car) away from the nearest city.

For the purposes of the HCS, we have collapsed Large Rural Towns with Small Towns/Isolated Rural Areas to form a single "Rural" category. The RUCA system applies to 2000 census tract boundaries (they have not been updated for 2010). Given this limitation, we are mapping all of the Census 2010 data (for demographic characteristics) into the Census 2000 boundaries (there are standard GIS tools for doing this translation between Census 2000 and 2010 boundaries).

3. Race/Ethnicity

The HCS is being designed to specifically over sample communities with high proportions of African American residents and Hispanic/Latino residents – as these populations are at increased risk for childhood obesity.

Stratification of census tracts for oversampling these important minority populations is based on the following:

- A census tract will be stratified into a high-proportion of **African American** group if 30% or more of the total population in the census tract is African American*
- A census tract will be stratified into a high-proportion of **Hispanic/Latino** group if 30% or more of the total population in the census tract is Hispanic/Latino*
- A census tract will be stratified into the **Other** group if less than 30% of the population in the census tract is African American and less than 30% of the population in the census tract is Hispanic/Latino

The selection of 30 percent was based loosely on the following definition of a minority census tract:

> *Banking regulations define minority communities when it comes to prohibiting redlining and other discriminatory lending practices. According to 12 USCS § 4502 (29), [Title 12. Banks and Banking; Chapter 46. Government Sponsored Enterprises] the term minority census tract means "a census tract that has a minority population of at least 30 percent and a median family income of less than 100 percent of the area family median income.*

***** For census tracts with both 30% African American and 30% Hispanic/Latino; the tract is classified in the category with the higher proportion (e.g., if a census tract contains 43% African American and 34% Hispanic/Latino residents – it would be classified as a high-proportion African American census tract).

Additionally, it should be noted that some residents within a census tract may be both African American and Hispanic/Latino. These residents would contribute to the calculation of both percentages when stratifying a census tract. Thus, the sum of the percentages (African American, Hispanic/Latino and Other) may be greater than 100% for some census tracts when doing these calculations.

4. Income

   The US Department of Housing and Urban Development (HUD) designates Qualified Census Tracts (QCTs) for purposes of the Low-Income Housing Tax Credit (LIHTC) program. The LIHTC program is defined in Section 42 of the Internal Revenue Code of 1986. The LIHTC is a tax incentive intended to increase the availability of affordable rental housing.

   The LIHTC statute provides two criteria for QCT eligibility. A census tract must have either: 1) a poverty rate of at least 25 percent; or 2) 50 percent or more of its householders must have incomes below 60 percent of the area median household income. The area corresponds to a metropolitan or a non-metropolitan area. Further, the LIHTC statute requires that no more than 20 percent of the metropolitan area population reside within designated QCTs (This limit also applies collectively to the nonmetropolitan counties in each state). Thus, it is possible for a tract to meet one or both of the above criteria, but not be designated as a QCT.

   The HCS is adopting this definition (whether a census tract qualifies for the LIHTC) as an indicator of income/poverty for the study.

5. Pre-Selection Intensity rating of Community-Based Programs/Policies (High, Moderate, and Low/ None – based on a scoring procedure as described below)

   Once the GIS database described in Section B.1.b.1 is populated, we will develop census tract level pre-selection scores to serve as the basis for stratification. The score will hierarchically integrate information across the different program and policy entries (which will occur at different levels of geographic specificity), with higher scores assigned to census tracts that multiple program and policy entries, higher funding, longer duration, and recommendations from program/policy funders or sponsors or HCS steering committee members for their promising approaches.

   It should be noted that these scores will be utilized only for the *selection* of census tracts (communities) into the study. For analyses of the HCS's results, much more rigorous data collection to assess community-based programs and policies will occur within each community selected into the study.

   A score will be calculated for each U.S. census tract, and each census tract will be assigned to a high-, moderate-, or low-score group based on this score - with the top 10% highest scoring counties assigned to the high-score group, the next 30% to the moderate score group, and the lowest 60% assigned to the low score group. These percentages are approximate guides and may be changed following inspection of the distribution of scores across all U.S. census tracts. Our intent is to oversample communities with active programs and policies aimed at reducing childhood obesity, with approximately 65 communities being selected within each group (high-, moderate- , and low-scores).

Several counties throughout the U.S. will enter our NPBS with certainty due to population size. In some counties, we will select multiple communities to preserve homogeneous sampling weights across communities selected into the HCS. The 15 largest counties in the U.S. (based on 2010 U.S. Census data) will be included in the HCS, along with the number of communities that will be selected within each of these large population centers. Table B.1.1. illustrates the 15 largest U.S. counties and the number of communities that will be selected within each county.

Table B.1.1 The Largest 15 Counties in the United States (as per 2010 U.S. Census)

County	State	2010 Total Population	Number of Communities
Los Angeles County	California	9,818,605	5
Cook County	Illinois	5,194,675	3
Harris County	Texas	4,092,459	3
Maricopa County	Arizona	3,817,117	2
San Diego County	California	3,095,313	2
Orange County	California	3,010,232	2
Kings County	New York	2,504,700	2
Miami-Dade County	Florida	2,496,435	2
Dallas County	Texas	2,368,139	2
Queens County	New York	2,230,722	2
Riverside County	California	2,189,641	1
San Bernardino County	California	2,035,210	1
Clark County	Nevada	1,951,269	1
King County	Washington	1,931,249	1
Wayne County	Michigan	1,820,584	1
Total		48,556,350	30

These 15 largest counties capture approximately 16% of the U.S. population and these counties account for 30 communities; therefore, 30 of the 195 communities within the NPBS will be proportionally assigned within these 15 counties as certainty counties within the NPBS.

We will follow the above described stratified sampling process of selecting the remaining 165 counties. It should be noted that the number of levels of each of the stratification variables [region (5), urbanicity (3), race/ethnicity (3), income (2), and pre-selection score (3)] yield 270 unique combinations. However, there are many of these unique combinations that have either zero or very few census tracts (and children). The sampling statisticians for the HCS have developed an approach for combining these unique combinations in a manner that will allow the HCS to assess difference in the association between community-based programs and policies and childhood obesity outcomes among communities with different race/ethnic profiles (e.g., comparing predominantly African American communities with predominantly Hispanic/Latino communities), and among communities with different urbanicities (e.g., comparing urban with rural).

Once the 165 census tracts are selected from each of the strata described above, as well as the 30 census tracts from the 15 large-population counties, we will identify the high-school catchment area associated with each selected census tract. The HCS will purchase the QED Education data-source from MCH (http://www.mchdata.com/), which will help us identify the public high school(s) that are in closest proximity to the selected census tract, as well as contact information for planners in the corresponding school districts. GIS researchers will then work to establish the current boundaries of the high-school catchment areas to complete the design, using an analysis of population at the census block level to identify which high school catchment area to select when a census tract is split among multiple high school catchment areas.

B.1.b.3 Selection of Certainty Communities

Following selection of the 195 NPBS communities, a multi-step process will be employed by a small (5-7 member) certainty community selection committee (CCSC) to select the 80 certainty communities. CCSC members will include experts with broad knowledge of community-based programs and policies aimed at reducing childhood obesity across the U.S.

Selection of certainty communities will utilize a stratified approach to ensure representation across the same factors used in the NBPS (Section B.1.b.1), including region (5 levels), urbanicity (3 levels), income (2 levels), and race/ethnicity (3 levels). We will designate combinations of the above four factors as a certainty selection strata (following a similar approach to combining/collapsing combinations with very few census tracts) – with a goal of selecting at least one certainty community within each of these established strata. Thus, the majority of strata will include only a single certainty community, with a minority of strata containing two or more certainty communities.

There are three main steps to select certainty communities:

1. The CCSC will select 80 communities based solely on Steering Committee nominations. These communities will be placed into the above strata, and the CCSC will provide an independent rating of each nominated community, using a 0-10 scale with 10 representing a community with the most highly promising programs and policies. The scores will be averaged across CCSC members for each community, and the ranked scores will be used by the CCSC to select communities.

   We anticipate that most of the strata will include one or more nominated communities; however, if there are strata that do not contain any nominated communities, the CCSC may ask the HCS Steering Committee to identify potential certainty communities in the empty strata.

2. Battelle will generate three rank-ordered lists of communities identified within the GIS database (described in Section B.1.b.1) with:
   a. the highest per-capita funding across all programs/policies that impact the community over the past decade (population size defined by U.S. Census 2010 Population estimates for the number of children aged 3-15 years);
   b. the highest number of programs and policies identified in the community over the past decade; and
   c. the highest number of independent funders/sponsors of the community programs/policies identified.

   These three lists, containing the top 100 ranked communities nationwide, will be categorized into the strata. The CCSC will review these rank-ordered lists to determine whether some communities on this list should be substituted for any of the 80 previously identified communities (in step 1). This step should ensure that most communities with promising programs and policies are considered even if not nominated by a Steering Committee member.

3. The CCSC will review communities (anticipated 150 – 200 communities) nominated through the interviews of staff at state health departments. These interviews will be used to identify communities with promising programs/policies that may not have been identified by the HCS Steering Committee or other federal and private sponsors, or that may be self-funding their activities. To facilitate review, these nominated communities will be placed into the strata, and the CCSC members will be asked to rate them using the same 0-10 scale. The scores will be averaged and communities will be ranked. The CCSC will substitute up to 20 communities from this review in place of previously selected communities, to ensure diversity among the communities selected with certainty.

In most cases, the certainty community selected by the CCSC will not match our definition of a community, i.e., it will cover a geographical area that contains multiple public high-school catchment areas. In these instances, we will identify all public high-school catchment areas within the selected area and select one with probability proportional to size (where the measure of size represents the 2010 U.S. Census population of children aged 3-15 years). These selection probabilities will be taken into consideration while developing the sampling weights at the community level, as described in Section B.1.b.4 below.

B.1.b.4 Calculation of Sampling Weights at the Community Level

Sampling weights at the community level will be calculated to represent the MOS corresponding to the strata from which the community was selected, where the MOS is based on the number of children aged 3-15 years from 2010 U.S. Census Data. For certainty communities, the sampling weight will represent the MOS within the geographic area originally identified by the CCSC. For communities selected through the National Probability-Based Sample, the selection weights will represent the MOS among all counties that were represented within the stratum after subtracting the MOS from any certainty communities that are located within counties belonging to that stratum. For the subset of 15 certainty counties that include multiple communities as detailed in Section B.1.b.2, each community selected will be assigned a proportional part of the MOS of the certainty county. This procedure ensures that the sum of weights across all communities selected in Wave 2 will be equal to the total population of children in the U.S. aged 3-15 years.

To facilitate the calculation of weights at the study participant level, the MOS for each community will also be computed for the 26 specific age-by-gender combinations.

B.1.b.5 Selection of Repeat In-Person Assessment (RIPA) Communities

The 40 RIPA communities will be selected as a stratified random sample of the 275 Wave 2 communities. The RIPA communities will represent the diversity of the Wave 2 communities in terms of population demographics and geographic location. This will ensure that the data collected in the RIPA assessments can be properly used in the statistical error and bias adjustments. These methods require an assumption that the more detailed information is missing at random in the other 235 communities. For logistical reasons, the 40 RIPA communities will be the first communities visited during the Wave 2 baseline in-person data collection thereby allowing for a repeat of the original in-person data collection on the participants and the community three years following the baseline assessment.

A stratified approach will be used so that sampling weights for RIPA study participants can be calculated for the three-year follow-up in a manner that allows these data to be representative of the entire U.S. population. Due to the limited degrees of freedom, we cannot use all of the stratifying variables being employed to select the non-RIPA communities. Thus we propose using 40 RIPA strata defined within combinations of the same geographic and socioeconomic variables as before: region (5 levels), urbanicity (2 levels: Urban & Suburban vs. Rural), income (2 levels), and race/ethnicity (3 levels). Similar to the MOS discussion in Section B.1.b.4 above, the MOS for the RIPA assessments will be developed by summing up the MOS among all of the communities assigned to each of the 40 RIPA strata. These RIPA MOS weights will only be used for calculation of survey weights for analyses that integrate the three-year RIPA assessments.

B.1.b.6 Temporal Ordering of Field Operations

The HCS plans to assign randomly the year of assessment (Year 1, 2, or 3 of Wave 2 data collection) to the selected Wave 2 communities while ensuring that all communities selected within the same county (whether chosen as a certainty community or via probability-based selection) complete their baseline assessment within the same year of operation. This will preserve the ability to compare data from communities within the same county, and increase operational efficiency.

The HCS plans to assign randomly the year of assessment (Year 1, 2, or 3 of data collection) to the Wave 2 communities using a stratified approach once they are selected for the following reasons: 1) communities will receive differential amounts and timing of follow-up assessments by data collection year and 2)we want to ensure that the proper weights based on strata MOS can be created for inferences based on nationally representative estimates from the follow-up assessments. In addition, for operational efficiency, we want to ensure that all communities selected within the same county (whether chosen as a certainty community or via probability-based selection) complete their baseline assessment within the same data collection year.

We will utilize the same 40 strata used in selecting the RIPA communities to assign all 275 Wave 2 communities into Years 1, 2, and 3 of data collection. The 40 RIPA communities will be visited in Year 1 of data collection, along with any other selected communities that are within the same county as that RIPA community. The 40 RIPA strata will then be randomly ordered, and communities will be identified from each stratum sequentially within this ordering; for example, a second Year 1 community will be identified from each stratum (along with any other selected communities within the same county) until 100 communities are identified for the first year of Wave 2 data collection. This same process will be used to identify the 100 communities for the following year, with the remaining 75 communities designated for Year 3 of Wave 2 data collection.

Once the communities are identified and the baseline assessment is assigned to a data collection year, Battelle will work with its partners (EMSI and the Universities) to propose a specific sequencing of data collection for consideration by the HCS Steering Committee. The sequencing may attempt to increase operational feasibility and likelihood of outdoor activities for children by performing baseline assessments in Northern areas (e.g., Chicago) in the non-winter months and Southern areas (e.g. Houston) during months when the heat is not too oppressive.

B.1.c Respondent Universe – Child/Parent Participants

B.1.c.1 Random Sample of Participants within each Selected Community

While the selection of communities follows a complex hybrid design as described in Section B.1.b above, the sampling design of child/parent study participants is less complicated. In Wave 2, the majority of communities (235) will include 78 child/parent study participants uniformly selected over yearly age intervals (3-15 years) and gender. The 40 RIPA communities will include a 50% greater sample – 117 children uniformly selected over age and gender. The larger sample size in RIPA communities will ensure that a sufficient number of study participants will be available three years following the baseline assessment for the repeat in-person assessments. The sampling of child/parent participants will follow a two-stage selection process, as described below:

Selecting Households: At the first stage, we will generate a random sample of households in each community (public high-school catchment area) that are likely to contain children in the 3-15 year age range using an independent data source (*infoUSA*™). The sample size generated will be approximately 10-fold greater than the number of study participants -- 800 households per community (1,200 for RIPA communities). The *infoUSA* database is built directly from thousands of original public and proprietary sources, including 4,300 telephone directories, U.S. Census data, state and local public records, product registrations and surveys (self-reported), and property/realty records, such as property deeds, mail-order transactions, and other proprietary sources. To measure accuracy of data items, *infoUSA* makes more than 20 million phone calls each year to verify information contained in their databases. *InfoUSA* also updates their database to account for approximately 300,000 residential moves each week, ensuring up-to-date contact information. We will purchase contact information from *infoUSA* for a sample of households within each selected community that are targeted with respect to child age criteria, while also including appropriate contact information (address and phone number). The addresses of all purchased contacts will be geocoded and verified to be within the high-school catchment area, and the resulting list of verified contacts will be placed into a random order and then sequentially sampled in batches by Battelle's survey implementation group as described below in the second stage to select child/parent participants within the selected households. When a potential household either refuses to participate or is ruled ineligible for the study (e.g., no children, the age/gender quota is already filled, family does not meet residency requirements), the next household in the sequence will be contacted.

Sampling Child/Parent Participants within Households: Exclusion criteria for sampling children in the household include:

1. no more than one child participant per household/family;
2. children who are institutionalized or are not ambulatory;
3. children from families that plan to move away from the study community within one year of the initial in-person assessment; and
4. children who have lived in the community for less than a year.

The second stage of the selection process begins with an informational letter (see SSA Attachment 4) and a study brochure (see SSA Attachment 5) sent to the targeted household. The

letter explains the HCS and provides the household with a unique study code and password to be used in combination with a website address or a toll-free 1-800 number that will allow them to initiate the process of screening and enrollment into the study. For households that do not initiate screening through the website or 1-800 number, Battelle's telephone service center will attempt to contact the household for screening and enrollment. Battelle telephone interviewers will attempt up to five calls to the household, at different times of the day and on different days of the week, to reach the adult in the household. If the adult cannot be reached within the five calls, that household will be removed from the calling list. Once contact is made with a household adult over the phone, we will attempt to identify the age and gender of all children living within the household as well as willingness to participate in the HCS (both for the child and parent).

During the recruitment call, a trained telephone interviewer from Battelle will: (1) explain the HCS to the household adult member (preferably the parent or main caregiver of the resident children); (2) enumerate resident children by age and gender to identify eligibility; (3) assess willingness to participate; and (4) explain that the HCS will randomly select a single child from the household to participate in the study (in households that contain more than one eligible child). Note that the eligibility of children within a particular household will be based on answers to the screening questionnaire as well as whether the HCS still requires a child with matching age and gender characteristics for that particular community (e.g., a household could have one or more children that have age and gender characteristics that match cells in the above table in which the sampling goal has already been met for a particular community). As long as one or more children within the household are eligible for the study, the HCS enrollment information management system will randomly select a single child to participate in the study. Selection probabilities will be based on the number of children with matching age and gender characteristics that remain to be recruited and enrolled into the HCS at the time of the telephone screening and recruitment session.

In a typical community, the study will recruit six children (three males and three females) within each yearly age interval (3-15 years), totaling 78 children per community. The number of planned child/parent home assessments for the HCS is detailed in Table B.1.3 below. The parents from whom we will collect anthropometric data are the biologic mother or father, or both.

Table B.1.3 Number of Planned Child/Parent Assessments by Type of Community

Year of Data Collection	Type of Community Sample	Number of Communities	**Participants Per Community**			**Participants Across Communities**		
			Standard Protocol Only	Enhanced Protocol	Medical Records	Standard Protocol Only	Enhanced Protocol	Medical Records
	Wave 1	4	65	13	55	260	52	220
Year 1	RIPA - Baseline	40	97	20	82	3,880	800	3,280
	Regular - Baseline	60	65	13	55	3,900	780	3,300
Year 2	Regular - Baseline	100	65	13	55	6,500	1,300	5,500
Year 3	Regular	75	65	13	55	4,875	975	4,125
	Remote Follow-up	200	55/82*	Not Applicable		12,080	Not Applicable	
Year 4	RIPA – Follow-up	40	65	13	N.A.	2,600	520	N.A.
**Total Number of Participants						19,415	3,907	16,425
Total Number of In-Person Assessments (does not include remote follow-up)						22,015	4,427	

* It is estimated that 70% of original participants will complete the remote follow-up assessments in Year 3 of the Wave 2 data collection, which would include 55 child/parent participants in regular communities and 82 in RIPA communities. ,
**Please note that information contained in the burden table is for the first three years of data collection, while Table B.1.3 includes data for the entire data collection period.

Table B.1.4 (below) demonstrates the recruitment goal by age and gender for each type of community involved in the HCS. [Note that due to an odd number of children being recruited into the RIPA communities within each age group (n=9), we created two sets of RIPA recruitment goals (Types A and B) in order to maintain gender balance across participants in these communities.]

Table B.1.4 Recruitment Goals for Children by Age, Gender, and Type of Community

Age of Child Study Participant At Enrollment, in years	Non-RIPA Communities Wave 1 (n=4)/Wave 2 n=(235)		RIPA Communities Type A (n=20)		RIPA Communities Type B (n=20)	
	Females	Males	Females	Males	Females	Males
3	3	3	4	5	5	4
4	3	3	5	4	4	5
5	3	3	4	5	5	4
6	3	3	5	4	4	5
7	3	3	4	5	5	4
8	3	3	5	4	4	5
9	3	3	4	5	5	4
10	3	3	5	4	4	5
11	3	3	4	5	5	4
12	3	3	5	4	4	5
13	3	3	4	5	5	4
14	3	3	5	4	4	5
15	3	3	4	5	5	4

B.1.c.2 Selection of Participants for Enhanced Protocol

The HCS will include a random sample of study participants (one of six, or approximately 17%) in Wave 1 and Wave 2 that will participate in the Enhanced Protocol, which will involve more detailed nutritional and physical activity assessments.

B.1.c.3 Calculation of Sampling Weights at the Participant Level

Sampling weights will be calculated for each study participant in the HCS for each stage of data collection that the participant completes (baseline, remote follow-up, and RIPA) by dividing the age-by-gender MOS described in Section B.1.b.4 by the number of study participants that complete data collection within that same age-by-gender combination for that particular wave. Alternative community MOS weights associated with the 40 RIPA communities (for the three-year follow-up assessment), and the first 200 Wave 2 communities (for the remote assessments) are also discussed in Section B.1.b. above – and can be utilized for the creation of sampling weights for specific analyses that utilize these data.

B.1.d Respondent Universe – Key Informants

Community key informants, knowledgeable of programs and polices targeting childhood obesity in their communities, will be asked to provide information to document the evolution of these programs and policies. The first key informants contacted in a community will be identified by (1) use of the QED Education data-source from MCH (http://www.mchdata.com/), which will

help us identify contact information for personnel associated with public schools within the catchment area and corresponding district, and (2) web-based searches to identify people within the municipal government (e.g., a representative from parks and recreation, city/county health department), and other organizations within the community (e.g., local YMCA, Chamber of Commerce, United Way, local hospital, etc.). First responders to the key-informant interview described above will also be asked to participate in snowball sampling to identify other community stakeholders for the study. Approximately 20 potential key informants will be screened in each community to identify 10 to 15 key informants to be interviewed that represent a variety of community interests and who consent to take part in the study.

Key informants will likely include these types of individuals identified within several key settings/sectors:

a) Schools (e.g., Principal of High School, Principals of two (based upon table below) randomly selected Middle Schools, Principal of one randomly selected Elementary School, school health coordinator, parent-teacher organization);
b) Health Organizations/Coalitions (e.g., chair of active community coalition, administrator/staff of local health department, administrator/staff of local hospital);
c) Government (e.g., city manager/administrator, staff of parks and recreation, urban planner); and
d) Non-Profit/Community Organizations/Service Agencies (e.g., administrator/staff of United Way, youth organization, community foundation, neighborhood organization, Chamber of Commerce, faith organization).

Table B.1.5 below provides a listing of the planned priority (and alternate) key informants by setting, as well as an indication of whether or not we would expect to ask the key informant additional questions specific to the physical activity environment and/or the nutritional environment within the community.

Table B.1.5 Priority/Alternate Community Key Informants Targeted for the HCS

SETTING/SECTOR	PRIORITY/ALTERNATE KEY INFORMANTS	SUPPLEMENTAL QUESTIONS	
		Physical Activity	Nutrition
SCHOOLS	PRIORITY: Principal of one High School	Yes	Yes
	PRIORITY: Principal of 2 randomly selected Middle Schools	Yes	Yes
	PRIORITY: Principal of 2 randomly selected Elementary Schools	Yes	Yes
	ALTERNATE: School health coordinator	Yes	Yes
	ALTERNATE: PE coordinator	Yes	
	ALTERNATE: Food service coordinator		Yes
HEALTH ORGANIZATIONS/ COALITIONS	PRIORITY: Chair of active community coalition (if present)	Yes	Yes
	PRIORITY: Administrator/staff of local health department		Yes
	PRIORITY: Administrator/staff of local hospital		
GOVERNMENT	PRIORITY: City manager/administrator		
	PRIORITY: Head/ staff of parks and recreation	Yes	
	PRIORITY: Urban planner	Yes	
NON-PROFIT/	PRIORITY: Administrator/staff of United Way		

SETTING/SECTOR	PRIORITY/ALTERNATE KEY INFORMANTS	SUPPLEMENTAL QUESTIONS	
		Physical Activity	Nutrition
COMMUNITY ORGANIZATIONS/ SERVICE AGENCIES	PRIORITY: Director/staff of community foundation with health mission		
	ALTERNATE: Director/staff of child or youth organization		
	ALTERNATE: Leader of active neighborhood organization		
	ALTERNATE: Head of Chamber of Commerce		
	ALTERNATE: Leader of active faith community		

Identified key informants will then be contacted directly via an information letter (SSA Attachment 11), a study brochure (SSA Attachment 12), and/or telephone by a Battelle research staff member (a Battelle community liaison) who will describe the study, determine the key informant's eligibility, and request his/her participation. Once a potential key informant agrees to participate in the study, he/she will be requested to provide documentation on community programs and policies related to childhood obesity and an in-person or telephone structured interview will be scheduled.

At the time of the interview, the Battelle community liaison will explain the study, review the consent document, and answer any questions the key informant may have. Following this, the key informant will be asked to sign the informed consent form prior to the initiation of the structured interview. When the respondent is unable to participate in an in-person interview and completes a telephone interview instead, a verbal consent script will be read to the key informant by a Battelle community liaison before the interview begins. Battelle community liaisons will be trained in Human Subjects Research; therefore, they will know the guidelines regarding what qualifies as informed consent.

B.1.e Power Calculations

The inclusion of 275 communities in Wave 2 of the study will allow the HCS to observe a considerable degree of natural variation in approaches and intensity of effort in childhood obesity programs and policies. The involvement of 78 representative children within each community across the 3-15 year age range will provide our study with strong power to observe subtle community-level impacts on the childhood obesity outcomes.

Researchers may be interested in the association between different program/policy components and childhood obesity outcomes for all ages combined and within a particular age group using a variety of approaches. One approach would use all data measured when a subject's age was within the age group of interest regardless of the subject's age at the baseline assessment. For example, if a participant who was 15 years old at the baseline assessment was found to have medical records containing BMI information when he/she was four years old, those records could be included in an analysis focusing on the 3-4 year-old age group. A second approach would use all data ever measured on an individual whose baseline age fell within the age group of interest, enabling specific cohort analyses. For example, a study focusing on 3-4 year-olds would include all data collected on those children that were ages 3 or 4 at the baseline

assessment. Our simulation study to assess power addressed both of these age-specific approaches.

Table B.1.6 lists the effect size as a percentage change in BMI that could be detected for a one-unit change (i.e., change in intensity score from 0 to 1) in program intensity score, assuming a cross-sectional analysis. Our analyses indicate that a cross-sectional study design with 275 communities, not including the four Wave 1 communities, and about 78 children per community, spread evenly across the age range from three to 15 have adequate power to identify programs/policies that reduce BMI by approximately 3-4%. The simulations cover scenarios where only a small fraction of communities will have implemented the program/policy in question with high quality and intensity, as well as the scenario where program implementation varies uniformly across communities. Additional details on the study design and the types of inferences that can be supported by this data collection effort can be found in Attachment **2**, which focuses on the Statistical Power Calculations.

Table B.1.6 Cross-Sectional Power and Effect Size Results for BMI Response

Power	Alpha	Age group	Effect Size (% change in BMI that can be detected)	
			Uniform distribution of program intensity across communities	**Few communities implement at full intensity, most communities at low intensity**
0.8	0.05	All ages	3.23%	4.27%
		3-4	4.38%	5.84%
		5-6	4.42%	5.88%
		7-8	4.42%	5.86%
		9-10	4.42%	5.88%
		11-13	4.01%	5.34%
		14-15	4.32%	5.73%
0.9	0.05	All ages	3.75%	4.96%
		3-4	5.09%	6.78%
		5-6	5.13%	6.83%
		7-8	5.13%	6.81%
		9-10	5.13%	6.84%
		11-13	4.65%	6.20%
		14-15	5.02%	6.66%
0.9	0.01	All ages	4.48%	5.93%
		3-4	6.08%	8.12%
		5-6	6.14%	8.18%
		7-8	6.14%	8.16%
		9-10	6.14%	8.19%
		11-13	5.56%	7.42%
		14-15	6%	7.98%

Table B.1.7 lists the effect size as a percentage change in BMI that could be detected for a one-unit change (i.e., change in intensity score from 0 to 1) in program intensity, assuming a longitudinal analysis. Our longitudinal designs highlight the utility of the historical height and weight data that may be obtained via medical record abstraction. These designs have much higher power than our cross-sectional designs. With historic longitudinal data, the study designs

have more than 90 percent probability of being able to identify programs that reduce BMI by less than one percent, even if those programs/policies are only implemented well in a small number of communities.

Table B.1.7 Longitudinal Power and Effect Size Results for BMI Response

Power	Alpha	Age group	Effect Size (% change in BMI that can be detected)	
			Uniform distribution of program intensity across communities	**Few communities implement at full intensity, most communities at low intensity**
Age group refers to age when measurement recorded				
0.8	0.05	All ages	0.37%	0.39%
		3-4	0.76%	0.83%
		5-6	0.92%	0.99%
		7-8	0.98%	1.07%
		9-10	1.07%	1.21%
		11-13	0.74%	0.87%
		14-15	2.70%	3.28%
0.9	0.05	All ages	0.43%	0.46%
		3-4	0.88%	0.96%
		5-6	1.06%	1.14%
		7-8	1.13%	1.23%
		9-10	1.23%	1.40%
		11-13	0.86%	1.01%
		14-15	3.13%	3.80%
0.9	0.01	All ages	0.51%	0.54%
		3-4	1.05%	1.14%
		5-6	1.26%	1.36%
		7-8	1.35%	1.47%
		9-10	1.47%	1.67%
		11-13	1.02%	1.20%
		14-15	3.74%	4.54%
Age group refers to age at baseline				
0.8	0.05	All ages	0.26%	0.27%
		3-4	0.31%	0.36%
		5-6	0.39%	0.42%
		7-8	0.54%	0.57%
		9-10	0.69%	0.71%
		11-13	0.64%	0.65%
		14-15	0.78%	0.79%
0.9	0.05	All ages	0.30%	0.32%
		3-4	0.36%	0.42%
		5-6	0.45%	0.49%
		7-8	0.62%	0.66%
		9-10	0.80%	0.82%
		11-13	0.74%	0.75%
		14-15	0.90%	0.92%
0.9	0.01	All ages	0.36%	0.38%
		3-4	0.43%	0.50%
		5-6	0.53%	0.59%
		7-8	0.74%	0.78%

Power	Alpha	Age group	Effect Size (% change in BMI that can be detected)	
			Uniform distribution of program intensity across communities	Few communities implement at full intensity, most communities at low intensity
Age group refers to age when measurement recorded				
		9-10	0.95%	0.98%
		11-13	0.88%	0.89%
		14-15	1.07%	1.09%

We also examined the power design to detect the impact that community programs/policies have on a binary outcome measuring physical activity or nutrition. Table B.1.8 lists the effect size as a percentage change in the probability that a participant has the binary outcome that could be detected for a one-unit change in program intensity, assuming a cross-sectional analysis. The results provided in Table B.1.8 demonstrate that the study will be able to detect less subtle associations between community program/policy intensity scores and the binary physical activity and/or nutritional outcomes measured at baseline (using cross-sectional analyses); the study will be well powered to detect changes of 4-6% attributable to the 1-unit change in community program/policy intensity score across all age-groups combined, and changes of 8-12% across most specific age groups.

Table B.1.8 Cross-Sectional Power and Effect Size Results for Binary Response

Power	Alpha	Age group	Effect Size (% Change in a Binary Physical Activity or Nutritional Outcome that can be detected)	
			Uniform distribution of program intensity across communities	Few communities implement at full intensity, most communities at low intensity
0.8	0.05	All ages	3.36%	4.46%
		3-4	8.40%	11.06%
		5-6	8.38%	11.03%
		7-8	8.38%	11.07%
		9-10	8.39%	11.05%
		11-13	6.88%	9.09%
		14-15	8.38%	11.08%
0.9	0.05	All ages	3.89%	5.15%
		3-4	9.69%	12.72%
		5-6	9.67%	12.70%
		7-8	9.67%	12.74%
		9-10	9.68%	12.72%
		11-13	7.94%	10.48%
		14-15	9.67%	12.75%
0.9	0.01	All ages	4.62%	6.12%
		3-4	11.47%	15.01%
		5-6	11.45%	14.97%
		7-8	11.44%	15.03%
		9-10	11.45%	15%
		11-13	9.42%	12.39%
		14-15	11.45%	15.04%

Table B.1.9 lists the effect size as a percentage change in the probability that a participant has the binary outcome that could be detected for a one-unit change in program intensity, assuming a longitudinal analysis. The results provided in Table B.1.9 demonstrate reasonably subtle associations can be detected based on longitudinal analyses of binary physical activity or nutritional outcomes from the Standard Protocol (collected at in-person assessments and planned remote follow-ups) – with the study being well powered to detect changes of 3-4% in these physical activity/nutritional outcomes attributable to the 1-unit change in community program/policy intensity score across all age-groups combined, and changes of 6-10% across most specific age groups.

Table B.1.9 Longitudinal Power and Effect Size Results for Binary Response

Power	Alpha	Age group	Effect Size (% Change in a Binary Physical Activity or Nutritional Outcome that can be detected)	
			Uniform distribution of program intensity across communities	**Few communities implement at full intensity, most communities at low intensity**
Age group refers to age when measurement recorded				
0.8	0.05	All ages	2.47%	3.16%
		3-4	7.72%	10.13%
		5-6	6.09%	7.87%
		7-8	5.90%	7.55%
		9-10	5.90%	7.55%
		11-13	4.83%	6.19%
		14-15	5.88%	7.55%
0.9	0.05	All ages	2.85%	3.66%
		3-4	8.91%	11.67%
		5-6	7.04%	9.08%
		7-8	6.81%	8.71%
		9-10	6.81%	8.71%
		11-13	5.59%	7.15%
		14-15	6.80%	8.72%
0.9	0.01	All ages	3.39%	4.35%
		3-4	10.56%	13.78%
		5-6	8.35%	10.75%
		7-8	8.09%	10.32%
		9-10	8.09%	10.33%
		11-13	6.64%	8.49%
		14-15	8.07%	10.33%
Age group refers to age at baseline				
0.8	0.05	All ages	2.38%	3.04%
		3-4	5.91%	7.58%
		5-6	5.91%	7.57%
		7-8	5.91%	7.58%
		9-10	5.91%	7.57%
		11-13	4.84%	6.21%
		14-15	5.90%	7.57%
0.9	0.05	All ages	2.75%	3.51%

Power	Alpha	Age group	Effect Size (% Change in a Binary Physical Activity or Nutritional Outcome that can be detected)	
			Uniform distribution of program intensity across communities	Few communities implement at full intensity, most communities at low intensity
Age group refers to age when measurement recorded				
		3-4	6.82%	8.74%
		5-6	6.83%	8.73%
		7-8	6.83%	8.75%
		9-10	6.83%	8.73%
		11-13	5.59%	7.17%
		14-15	6.82%	8.74%
0.9	0.01	All ages	3.27%	4.18%
		3-4	8.10%	10.36%
		5-6	8.10%	10.35%
		7-8	8.10%	10.37%
		9-10	8.11%	10.35%
		11-13	6.64%	8.51%
		14-15	8.09%	10.36%

Based on a number of assumptions, power for this study was estimated using Monte Carlo simulations in R (version 2.12.0). Multiple datasets were simulated and analyzed using the mixed model routines in the lme4 package. The standard error of the program effectiveness parameter was recorded after analyzing each simulated dataset. The average of the standard error over all datasets was used to relate power to effect size using the following equation:

$$\beta_1 = SE \times \left(z_P + z_{1-\alpha/2} \right)$$

where SE is the average standard error estimated from the simulations, P is the power, α is the significance level of the test, and z_u is the $100*u^{th}$ percentile from a standard normal distribution.

In the cross-sectional power studies, we expect to utilize only the BMI data collected during the baseline home visit. There are no data available to examine changes over time. For the longitudinal study design, we expect 70 percent of the children to have medical records available that contain height and weight information. The number and timing of the previous medical records were simulated from a Poisson process assuming that the average number of medical visits was two per year up until age three, one per year from ages three to 10, and one every two years after age 10. The pattern of medical visits was allowed to be interrupted by a randomly drawn change in physician, since the study's request for medical records will be sent to only one physician per child participant. For each simulated participant, we selected the longest sequence of visits between any changes in physician as the data values to be analyzed.

Each child was assumed to receive a baseline assessment of the binary measure of physical activity or nutrition. For the cross-sectional analysis, this would be the only data point available. For the longitudinal study design, we assumed that 70% of the children in each of the first 200 Wave 2 communities would have a follow-up assessment as planned, the timing and number of which were simulated exactly as specified in the proposed study design. For children in the

RIPA communities, follow-up assessments occurred two years and three years after the baseline assessment. For children in non-RIPA communities sampled during the first year of Wave 2 data collection, a single follow-up assessment occurred two years after the baseline assessment. For children in communities sampled in the second year, a single follow-up assessment occurred one year after the baseline assessment. Children in the 75 communities sampled in the third year of Wave 2 data collection did not receive any follow-up assessment.

For the longitudinal analyses, we needed to develop a model that explains the change in program/policy intensity over time. Program/policy intensity was assumed to follow a logistic curve with a slope that could be positive (allowing program intensity to increase over time) or negative (allowing program/policy intensity to decrease over time). Each community's slope was randomly sampled from a normal distribution with a mean of zero and a standard deviation of 1.25.

B.2 Procedures for the Collection of Information

Data collection within the selected communities will be hierarchical by nature, and will include two primary components aimed at (1) documenting how childhood obesity programs and policies have evolved within the community over the previous ten years, and (2) assessing current anthropometrics (height, weight, BMI), physical activity, dietary behaviors, a longitudinal history of BMI abstracted from medical records, and a variety of other important covariates and confounders on a representative sample of children from within the community. This data collection will occur over a 3.5-year period. Figure B.2.1 provides an overview of data collection activities, with additional detail provided in the supporting documentation.

Figure B.2.1 HCS Primary Data Collection Activities in Wave 2, 2012-2015

Baseline assessment includes all home assessments above *plus*

B.2.a Community-Level Data Collection

Data collection at the community level will include a retrospective and prospective review of programs and policies to (1) document how childhood obesity programs and policies have evolved within the community over the previous ten years, and (2) how these programs continue to evolve during our study period. Documentation of the evolution of programs and policies within communities will occur via key informant interviews using multiple modes of data collection (telephone interviews, web-based questionnaires, in-person meetings). In addition, these interviews will be supplemented by document retrieval and abstraction by research staff. Please refer to SSA Attachment 14 for the key informant interview protocol.

Data collection at the community level will also include assessments of the broader community, as described in Section B.2.a.3 below.

B.2.a.1 Document Review

One component of this data collection will be the initial gathering and comprehensive review and documentation of program and policy-related information. During our baseline key informant interviews in all communities, a Battelle community liaison – a full-time, trained research staff member - will request to review documents developed by or about relevant community programs and policies. The types of documents we will review include: publicly available legislative hearing documents; annual program reports; management information system (MIS) reports often used by funders; available and relevant Requests for Proposals (RFPs) and responses to RFPs that may have been prepared within the time frame of interest; agency-wide reports such as Healthy People 2000 and 2010; coalition reports from relevant organizations, schools, and other institutional and community wellness policies.

Using a field-tested document abstraction instrument, data will be reviewed and then entered into an electronic form. This form can then be easily updated during our prospective program/policy review to be conducted in our RIPA communities (as described later).

B.2.a.2 Key Informant Interviews

In each community, the Battelle community liaison will identify and conduct interviews with key informants associated with and/or knowledgeable of relevant community programs and policies. We will use standardized data collection instruments to gather characteristics for each program and policy operating in the selected communities, such as: target population, including age, sex, location; target focus area; who delivered the program/services; level of funding and how it was funded; date initiated and ended, when applicable; number reached, and who was reached (i.e., providers trained and recipients of service); nature of collaborations; and any outcomes examined.

Key informants will answer a similar set of interview questions for each specific local program and/or policy identified during key informant interviews. Data will be combined from different key informant interviews for the same program or policy.

Remote Follow-up among Key Informants: Key informants that were interviewed during the initial baseline assessment in the first 200 Wave 2 communities will be re-contacted during the follow-up period (in Years 2 and 3 of data collection) when we re-contact children and parents for remote data collection. At that time, these key informants will be asked to complete about 50 minutes of data collection via a telephone survey. Key informants in the RIPA communities will be visited in-person a second time three years later to repeat the baseline assessment. Those key informants in the RIPA communities that could not be visited face-to-face will be asked to answer questions by web or telephone (similar to the original baseline assessment).

Identifying/Recruiting Local Community Documenters in RIPA Communities: During the interviews of key informants within the RIPA communities, the Battelle community liaison will discuss the role of the local community documenter and ask the respondent(s) to recommend potential community experts to play this role as part of the HCS. The recruited expert should have broad-based knowledge of the programs and policies that have been implemented within the community, as well as a willingness to perform the quarterly assessments and other planned assessments of the physical activity and nutritional environment within the community. The

local community documenter could be one of the key informants included in the study as a participant. Community documenters will be paid a yearly stipend.

B.2.a.3 Other Community Assessments

Community assessments will include interviews with additional school administrators/personnel, community key informants, participant perceptions of the school and home environments, GIS data, and direct observations of communities to collect program/policy and environmental data. The following assessments will be conducted in all communities for baseline and in selected communities for follow up.

Nutritional Environment: Baseline observational assessments of the nutritional environment will be conducted in up to four randomly selected schools (two elementary and two middle schools) per community. A member of the school's food service staff will complete a brief self-administered questionnaire, and the Battelle community liaison will observe the school's lunch period and complete an observation form. In addition, questions will be asked of the principals of the schools selected for observations during the key informant interview.

Elementary and middle schools were selected for assessment because, given the age range of the subjects of this study, the highest exposure will be to the elementary and middle school environments. Relatively few subjects will have had exposure to the high school environments.

Both the food service staff and the Battelle community liaison will provide information regarding:

- school characteristics -- meal program participation, free and reduced price meal eligibility, open campus status, participation in key federal school nutrition and food programs, degree and extent of scratch food preparation, extent of wellness policy implementation and how long in place, if and when changes have been made to the school meals, meal facilities and competitive foods;
- characteristics of reimbursable lunch options -- number and type of entrees, beverages, whole grain vs. refined grain products, fruits, vegetables, dessert, and snack foods;
- characteristics of dining facilities -- availability, ambiance, size;
- cafeteria staff interactions with students;
- meal service line length;
- time to eat;
- water availability; and
- competitive foods -- what, where, and how much is offered.

These data will be used to characterize the past ten years of the school food environment and validate and add detail to stakeholder reports of school-based efforts. Furthermore, the data will be used for cross-sectional, retrospective, and prospective longitudinal analyses of associations with anthropometric and behavioral measures.

As previously mentioned, the key informants will undergo follow-up interviews (remotely in the first 200 Wave 2 communities and then again in person in RIPA communities). The principals of the schools that were randomly selected for observations, if they consented to be a key

informant, will be sent a link to the food service staff questionnaire during the follow-up period in Year 3 and asked to forward the link to the appropriate food service person for completion.

GIS mapping and analyses will be conducted to characterize the community food environment at baseline and retrospectively 10 years prior to coincide to the extent possible with baseline and endpoint BMI data collection. The GIS data will include the density of various types of retail food establishments in the target community as well as their proximity to schools in the target area to characterize the community nutrition environment. In the RIPA communities, the community liaison will return to the community within three months of their baseline visit to conduct limited ground-truthing of the GIS data that were collected. For example, for the nutrition environment, this may entail verifying the existence of commercial food venues identified in the GIS data.

Physical Activity Environment: Baseline observational assessments of the physical activity environment will be conducted in up to four randomly selected schools (two elementary and two middle schools) per community. The Battelle community liaison will interview a member of the physical education staff at the school, and will also observe the school's physical activity resources using the Physical Activity Resource Assessment (School PARA) form. In addition, questions will be asked of the principals of the schools selected for observations during the key informant interview.

Furthermore, key informants who are particularly knowledgeable about parks and recreation and other physical activity resources in the community will answer additional questions. The Battelle community liaison will conduct a 15 minute interview with this individual to gather information about community resource availability and accessibility, physical activity related community collaborations, park and trail use and general features, and other information related to physical activity resources in the community.

The physical education interview will gather data about physical activity resources and facilities on school campuses, the provision of physical education, recess, and other physical activity opportunities at schools, the community partnerships established for providing physical activity opportunities, and the norms and culture for physical activity at the schools.

The School PARA will be conducted for both indoor and outdoor features of the environment related to physical activity. This form characterizes the features, amenities, and incivilities of the physical activity environment(s) in the school, the hours, availability, and capacity of the facilities, and size and cost of use of these environments.

GIS mapping and analyses conducted at baseline will also provide information related to the physical activity environment. When the Battelle community liaison returns to the community within three months of their baseline visit in the RIPA communities, s/he will also conduct limited ground-truthing of the physical activity GIS data. For example, for the physical activity environment, this may entail verifying the existence of parks. The physical activity environment will be further characterized in the RIPA communities by the completion of the full PARA for schools, parks, and trails identified within the community.

B.2.a.3 Prospective Assessments in RIPA Communities

In RIPA communities, the key informants who were interviewed at baseline will be contacted to participate in a follow-up interview with the Battelle community liaison. This interview, in addition to a full program documentation review and abstraction, will serve to document the disposition and outcomes of the various programs and policies discussed at baseline (approximately three years prior), which can then be linked to any noted changes in child participant health outcomes (mainly BMI) in that community.

Contact with key informants in the RIPA communities will also occur through designated community documenters (as described in Section B.2.a.2). On an approximate quarterly basis, following the baseline interview, the community documenters will contact key informants to access and then review and document any new or updated program and policy literature. Upon review and abstraction, the community documenter will update the electronic program/policy data abstraction form accordingly.

B.2.b Data Collected at the Child/Parent Participant Level

B.2.b.1 Anthropometrics

Anthropometric measurements will be taken at the initial household visit for all participants (and repeated three years after baseline for participants in the RIPA communities). Anthropometric measurements to be collected include the height, weight, and waist circumference of the child and the height and weight of the biological parent(s) either via measurement (if available during home visit), or self- or proxy-report. Measurements will be made according to the NHANES protocol, recorded in metric units (centimeters and kilograms), and measured to the nearest .1 cm and .1 kg. BMI will be calculated by dividing weight in kilograms by height in meters squared (kg/ m^2). Height and weight will be measured twice during each home visit. For participants who receive the Enhanced Protocol, anthropometry will be measured one week later in addition to during the first home visit.

B.2.b.2 Medical Record Review

Medical records for each child participant will be abstracted to develop, where possible, longitudinal BMI trajectories for the previous 10 years. Any indication of nutritional, physical activity, or sedentary activity counseling will also be abstracted from medical records. The presence of other chronic conditions and prescribed medications for those conditions (e.g. asthma, diabetes) will also be abstracted. At the conclusion of each community assessment, our subcontractor, EMSI, will contact one medical provider for each participant who provided consent to access medical records. Providers will be selected based on being the provider that the child has seen most often. The EMSI data abstractor will load the information into an electronic form that they will upload to the secure study data repository at the conclusion of the abstraction process. Based on previous experience, we expect that 70% of parents will consent to review of their child's medical record.

B.2.b.3 Nutritional Assessments

Standard Protocol: Information on food and beverage intake, food patterns and behaviors, breastfeeding, household food insecurity, and perceived family social support, perceived home environment, and perceived community environment regarding healthy eating will be collected

on children of all ages enrolled in the study. For children six years or older, questions will also be asked about the perceived school environment regarding healthy eating. For children 12 years and older, additional questions will be asked regarding self-efficacy, motivation, dieting behaviors, and perceived social support from friends related to healthy eating. Questions will be age appropriate, and either be self-administered (child and parent questions) or parent assisted, depending on the age of the child.

Enhanced Protocol: In addition to the standard nutritional assessments, participants selected for the Enhanced Protocol will complete two 24-hour dietary recalls approximately one week apart using the ASA24 (Automated Self-Administered 24-hour Recall). This online web-based 24-hour dietary recall is a modified version of the U. S. Department of Agriculture (USDA)'s Automated Multiple Pass Method (AMPM), which utilizes multi-level food probes to obtain estimates of types and amounts of food consumed on the day prior to the visit. The ASA 24 has been developed for self-administration, will be completed at the first and second home visits for Enhanced Protocol participants at baseline (and will again be administered twice during the in-person follow-up in the RIPA communities for Enhanced Protocol participants). Data from the 24-hour dietary recall will provide estimates of total intake for all nutrients available in the Food and Nutrient Database for Dietary Studies (FNDDS), food groups included in the Food Equivalents Database (MPED) and selected other food sub-groups such as sugar-sweetened beverages as well as sufficient information to derive summary scores of dietary quality, such as the Healthy Eating Index-2005 (HEI-2005). The purpose of collecting two recalls on each child is to enable adjustment for within person variability. The adjusted food and nutrient estimates from the ASA24 will be used to calibrate estimates on children in the entire sample obtained from the Household Interview dietary screener.

B.2.b.4 Physical Activity and Sedentary Behavior

Standard Protocol: All participants will answer questions related to their physical activity (e.g., types of activities; including intensity, frequency, and duration) at home, at school, and in the community, *during the previous week,* and their parents will be asked questions related to their children's activities as well as physical activity resources available at home and in their community. Questions will be gender specific and age appropriate, and either be self-administered (child and parent questions) or parent assisted, depending on the age of the child.

Enhanced Protocol: In addition to the Standard Protocol, the child participants will be asked to recall and describe the activities they participated in at home, at school, and in the community, *the previous* day using a similar structured instrument based on gender and age. Participants will also be given an accelerometer at their first home visit (at baseline). Accelerometers provide an objective measure of physical activity by detecting movement and intensity of activity. The device will be worn by participants except while bathing, swimming, or sleeping, for the week in between home visits.

B.2.b.5 Demographics, Family Medical History, Direct Observations

All parent participants will answer the following demographic questions related to the child and parent: age, race, ethnicity, marital status, country of origin, education, language, employment, and family income. In addition, the child's biological parents' height and weight will be measured (or reported if biological parent refuses measurement).

A series of questions are included in the Standard Protocol, aimed at assessing the medical history for the child participant related to:

1. Issues related to the participant child's access to healthcare;
2. Medical conditions that may alter diet for the participant child (diabetes, celiac disease, anorexia, bulimia, etc.);
3. Medical conditions that alter physical activity for the participant child (disabilities, recent accidents (e.g., broken ankle), etc.);
4. Medical conditions that alter ability for the participant to self-complete aspects of protocol (cognitive deficits, Down's syndrome, dyslexia, etc.).

B.2.b.6 Direct Observations of the Child's/Family's Neighborhood

In all communities, direct observations of the child's home will be completed by study staff. For children who receive the Standard Protocol, EMSI field interviewers will complete a five-item modified windshield survey during the baseline visit. EMSI field interviewers will rate features of the social and physical environment on the street segment associated with each child's home address. A street segment is defined as the street in front of the home, from intersection to intersection, not to exceed 0.5 miles. In the instances where the street segment exceeds 0.5 miles, the EMSI field interviewers will be instructed to consider the street segment that is contained within 0.25 miles from the home in either direction or to the nearest intersection, whichever is closer. A paper form of the modified five-item windshield survey will be completed when the EMSI field interviewer arrives at the child's home and entered into the study database following the home visit. This will be done at baseline and at follow up during the in person assessments in the RIPA communities. The modified five-item survey is expected to take less than five minutes to complete. GPS coordinates will also be obtained for each participant's household to match to GIS data.

B.3 Methods for Maximizing Response Rates

All data regarding contact success, refusal to screen and/or enumerate household members, refusal to participate for one or more eligible children (including reasons for non-participation) will be maintained to calculate response rates for the study. Response rates will be calculated as the fraction of eligible participants who elect to complete the data collection protocol, and will be calculated for child/parent participants as well as key informants.

Every effort will be made to maximize the response rates for key informant interviews, parent and child surveys, as well as acquisition of medical records by the participating child's medical provider. Based on our review of the current literature as well as our experience in the field with prior studies, the approaches that we will use to maximize the response rates, are described below.

Key Informant Interviews: Community key informants include, for example, high school principals, directors of programs involved with children health issues, and leaders of local government programs and non-profit community organizations. It is anticipated that participation will be relatively high among this group given their a priori interest in helping improve the health and well-being of those living in their community. Furthermore, as this

group will be comprised through a snowball sampling technique, we expect that the referral by one community leader to another will improve succeeding participation rates. To maximize response rates among this group, upon scheduling the initial appointment, we will maintain contact with a confirmation letter, including printed information about our study (see SSA Attachment 12), and follow-up confirmation telephone calls. Our protocol calls for conducting interviews with key informants in person, which is likely to improve our response rate, although when necessary due to scheduling or other conflicts, we will complete interviews over the telephone. All interviews will be conducted by the Battelle community liaison – a qualified and well-trained interviewer, who is well versed on the community and its programs/policies, and who is comfortable speaking with high-level community leaders and able to adeptly answer any study related questions. As we will also be requesting specific program/policy documentation from this group, we will follow-up on this request through our confirmation letter, subsequent telephone calls, and during our in-person interview meeting as needed.

Parent/Child Surveys: Battelle Community Liaison's will post study-related informational flyers and/or posters in locations most likely to be seen by potential respondents (e.g., within the lobbies of local schools, places of worship, social service agencies) when they first enter a community. In so doing, we hope to have information about the study seen by potential parent/child participants, thus encouraging participation and overall response rate. Following these community wide efforts, a brochure (SSA Attachment 5) and an informational letter (SSA Attachment 4) will be sent to targeted households to explain the HCS. The letter will also provide the household with a unique study code and password, to be used in combination with a web address or a toll-free 1-800 number, which will allow them to initiate the process of screening and enrollment into the study. For households that do not initiate screening through the website or 1-800 number, Battelle's telephone service center will make repeated telephone contact (up to five) to attempt screening and enrollment. Once we make contact with a household adult over the telephone, we will identify the age and gender of all children living within the household and the willingness to participate in the HCS (for both the child and parent).

Completion of parent and child baseline surveys in the home administered by well-trained EMSI field interviewers, rather than having the parents respond to a survey mailed to their home, is likely to result in a high response rate. In addition, when necessary, interviews will be available to be conducted in Spanish. Similar to the NHANES protocol, should neither English nor Spanish be spoken, the household member completing the family recruitment process will be asked to identify a family member or neighbor, aged 18 or over, who can be present during the home visit to translate to the interview questions into the necessary language. This will help maximize our initial enrollment and follow-up response rates.Furthermore, the offering of an incentive, though small, to both participating parent and child, is anticipated to further improve the response rate.

The challenge will be achieving a high response rate to the parent/child remote follow-up surveys, to be completed in Year 3 of the Wave 2 data collection, and the follow-up interviews to be completed in the RIPA communities three years after the baseline assessment. The remote follow-up interviews will be conducted over the telephone or through a web-based survey (the ease of which will likely reduce participant burden and improve follow-up response rates), while

the follow-up interviews in the RIPA communities will be completed during an in-home visit. To minimize loss to follow-up, we will obtain contact information for each parent and child to be surveyed and will record the information on the participant enrollment form at the baseline interview. In addition, we will ask them to provide the name and telephone number of friends or family members who will always know where to contact them in the future, in case s/he has moved and we are having trouble contacting them. During the intervening period between baseline and follow-up, we will undertake cohort maintenance efforts, including sending periodic post cards with the study name and design elements, and study-related materials (e.g., newsletters), reminding parents and children of their continued enrollment in the study and our plan to contact them at the designated follow-up time (with bilingual mailings and e-mailings when appropriate). Mailings will be addressed to the parent and/or child, and it will stress the importance of their continued participation with the study. All mailings will also include a toll-free telephone number for participants to call if they have any questions or want to provide new contact information. Any items mailed will also include a "Return Service Requested" stamp to help us track families that may have moved. For those mailings that are undeliverable, we will call the contact numbers provided on the enrollment form in an effort to obtain a current address. Prior to making follow-up telephone calls, we will mail a postcard to the parent alerting him/her about an upcoming telephone contact by a member of our research staff to schedule a follow-up telephone (or, if appropriate, an in-home) survey; this contact mailing will also remind parents of the additional incentive they will receive for their follow-up participation. All of these efforts, likely will result in promoting high follow-up response rates.

Medical Provider Records: It is anticipated that the number of records to be requested of any one provider will be relatively few. In addition, because we have budgeted to reimburse any administrative fees the providers may charge for copying or providing these records, and because our partner, EMSI, has a great deal of experience obtaining patient medical records from providers, we anticipate our response rate to this particular data collection effort to be relatively high.

B.4 Tests of Procedures or Methods to be Undertaken

All data collection instruments will be tested in house by members of the research team to ensure that timing will align with the estimate of respondent burden. During Wave 1 of data collection, we will visit four communities to refine recruitment materials, data collection instruments, and all field procedures. The Wave 1 communities will be chosen near existing Battelle offices to allow for observation and quality control checks of all protocol methods. If any refinements to sampling and selection procedures, field protocols, or data collection instruments are made, we will inform OMB of all changes and provide updated and final documentation and instruments.

B.5 Training and QA/QC Methods

Experienced trainers from Battelle and its partners will oversee and/or conduct all staff training, including training of the Battelle community liaisons and community documenters on conducting face-to-face interviews, document abstraction and community observations. The Battelle community liaison will also be trained on how to train EMSI field interviewers to accurately use and record measurements from the various anthropometric measurement devices (e.g., scale,

accelerometer), and how to conduct quality assurance and control of EMSI field interviewers once they begin conducting home visits.

Training sessions will be conducted using a number of modalities, including in-person one-to-one and group sessions (particularly for the training on how to conduct face-to-face interviews with key informants), on-line individual coursework, video viewing, and live webinars. As part of our training and quality assurance protocol, all study staff will also receive a personal training and reference manual, and will complete human research ethics trainings and confidentiality trainings.

B.6 Consultations and the Project Team

Individuals Consulted on Statistical Aspects of the Design:

Warren Strauss
Battelle Memorial Institute
614-424-4275

Eric Leifer
National Heart, Lung, and Blood Institute
301-435-0419

Colin Wu
National Heart, Lung, and Blood Institute
301-435-0440

Kevin Dodd
National Cancer Institute
301-496-7461

Contractors Responsible for Collecting Information for the Agency:

Battelle Memorial Institute
505 King Avenue
Columbus OH, 43201-2693
Contact: Dr. Howard Fishbein
(703) 248-1647