Supporting Statement B for

Continuation of NEXT Generation Health Study – NICHD

[OMB No. 0925-0610]

November 5, 2012

Project Officer:

Dr. Ronald J. Iannotti
Prevention Research Branch
Division of Epidemiology, Statistics, and Prevention Research
National Institute of Child Health and Human Development
Building 6100, 7B05
9000 Rockville Pike
Bethesda, Maryland, 20892-7510
Telephone: (301) 435-6951
Fax: (301) 402-2084
E-mail: ri25j@nih.gov

Table of Contents

List of Attachments

Attachment 1A: Data Collection Instruments – Participant Survey Hard Copy

Attachment 1B: Data Collection Instruments – Participant Online Survey Screen Shots

Attachment 2A: Data Collection Instruments – Peer Survey Hard Copy

Attachment 2B: Data Collection Instruments – Peer Online Survey Screen Shots

Attachment 3: Variable Source Table

Attachment 4: Participant Informed Consent Form (NEXT study)

Attachment 5: Participant Informed Consent Form (NEXT Peer study)

Attachment 6: Parent Informed Consent Form (NEXT Peer study under 18)

Attachment 7: Data Collection Instruments – Physical Activity Diary

Attachment 8: Information Technology System Security Plan

Attachment 9: Certificate of Confidentiality

B. Collections of Information Employing Statistical Methods

B1. Respondent Universe and Sampling Methods

A primary goal of NEXT is to examine the prevalence and determinants of selected health behaviors and health status measures in a longitudinal study of a nationally representative probability sample of students in grade 10 from public and private schools. The design provides estimates of population percentages with a margin of error of plus or minus 3 percentage points at the 95% confidence level. An oversample of African American students (Hispanic students are already adequately represented in the sample) in grade 10 is also included in order to improve the validity of sub-group analyses and to better study health disparities. Toward these ends, we constructed a sampling frame that was current, complete, and accurate with respect to information needed for selection and stratification. The population of interest included all 10th-grade students in public, private, and parochial schools in the 50 states and the District of Columbia. The sampling frame for the construction of primary sampling units (PSUs) and selection of public schools was the list of school districts supplied the Quality Educational Data, Inc. (QED). QED maintains a continuously updated list of every school district in the U.S. and is therefore current. It also maintains a current list of K-12 schools by state with contact information covering 100% of public, private and Catholic schools by State in the U.S. The list of school districts and schools has comprehensive data on enrollment by grade, race/ethnicity in addition to address and contact information. We had very few problems in terms of missing schools or misclassification by grade etc. in using this list for the selection of the sample and contacting selected schools based on the information provided on the list. These lists provided by QED required very little work in terms of adding information needed for building sampling

frames for selection of primary sampling units and schools. Hence we chose to use the lists supplied by the QED for NEXT.

The list of school districts was used to construct PSUs. PSUs were formed by grouping school districts within each Census division. The total number of PSUs created for selecting public schools was 1,302. For example, a primary sampling unit may contain all school districts within a county or two adjacent counties. Some PSUs contained only one very large school district. For sampling students from public schools, PSUs, which were either individual school districts or groups of school districts were created and selected as a sample of PSUs at the first stage. A data file that identified and provided extensive data on school districts and individual schools was purchased from The Quality Educational Data, Inc. (QED) and examined to construct a sampling frame of PSUs. The QED files were current and contained data on primary and secondary public schools as well as private and parochial schools. Private and parochial schools were linked to public districts to ensure that these sampled schools fell within the same sample clusters as sampled public schools.

While the sample of private and parochial schools was proportionately smaller than the sample of public schools, our recruitment rates for private schools were comparable to the recruitment rates of public schools. We used experienced recruiters and methods to explain the value of participating in the study and stressed the need to make sure that “the voices” of private and parochial schools and students were adequately represented in the national findings. This strategy was successful.

Sample Selection Procedures.

A primary sampling unit (PSU) was either an individual school district or a group of school districts in adjacent counties. These were created using a population of around 14,000 school districts. At the first stage of sampling, a sample of primary sampling units was selected. The list of schools offering grade 10 was obtained for only the selected primary sampling units. There was no sampling of school districts within a selected PSU as the list of schools was formed based on all the schools in the PSU without regard to school district. From this list, a sample of schools was selected. Only after the selection of the school, was the school district identified for purposes of contacting the school. This method of sampling reduced the cost of data collection as the sample of schools was not spread very widely across the U.S. Also, if we had wanted to directly sample schools from a list of schools, we would have needed a complete sampling frame of schools which is a list of all schools in the U.S. offering grade 10, where in a multi-stage design we only needed the list for selected school districts. It would have been more expensive to get a complete and correct list of schools offering grade 10 than just restricting the list to selected PSUs. We contacted a probability sample of 137 schools and 80 (58.4%) agreed to participate in the survey. We conducted response bias analysis to determine if the schools that consented to participate in the study were different than the schools that refused. The ONLY significant difference between schools that participated and those that refused was on the proportion of Asian American students. Because of the relatively small difference in the proportion of Asian American students in both groups (approximately 3%), this difference could have been due to the population of a single school in the refusal group and/or the oversample of schools with a high proportion of African American students.

For determining the sample size at the initial wave, a recruiting or retention rate of 85% at each wave (conservative compared to the previous work by us and others) was assumed as well as a response rate of 95% (conservative or consistent with previous work by us and others) from those students who were successfully recruited or retained in the sample at each wave. The required sample size at the end of wave 7 in terms of the number of completes was estimated based on the desired precision of the estimate of change between two time periods. This sample size should be such that we are able to reject the hypothesis of no difference in population percentages of characteristics of interest between two time periods (for example year 6 and year 7) with 80% power when actually there is a difference of 5.3 percentage points with a two-sided statistical test at 5% of level of significance. For the determination of the sample size we assumed that the correlation between two time periods was around 0.5. The sample was first determined assuming a simple random sample of students. This gives a sample of around 700 students. Since the sample is selected using a multi-stage sampling design, we assumed a design effect of 1.5 based on previous HBSC surveys and increased the sample to 1,050 completes in the main sample. The margin of error of the estimated population percentage at 95% confidence level at the end of wave 7 based on a sample size of 700 (or 1,050 with correction for the design effect) is plus or minus 3.7 percentage points.

The strategy for minority oversampling was based on the requirement of around 215 African-American students at the end of wave 4 out of sample of 1,050 completes. We expected to get around 180 African American students at the end of wave 7. Therefore, there was insufficient sampling of minorities in the basic sample. To get the additional minority students, we identified school with a high percentage of African American students and selected

additional samples of students to screen and identify minority students. Originally it was planned to select additional primary sampling units for sampling Hispanic students. This plan was not necessary. We were able to recruit the required number of 215 Hispanic students without oversampling as the percentage of Hispanic students was slightly higher than African-American students.

Of 3,796 students recruited to participate, youth assent and parental consent were obtained from 2,619 (69.0%) students. Of those consented, 2,524 (96.4%) completed the Wave 1 survey. In Wave 1, confidential self-report surveys were administered by trained research assistants in the 10th-grade classrooms.

The retention rate from Wave 1 to Wave 2 was 87.1%. Due to a delay in obtaining final approval from the Chicago Public Schools, 246 students from Chicago Public Schools did not participate in the study until Wave 2. As a result, the Wave 2 sample was 2,454. We have completed Wave 3 and a conservative estimate of number of completes in future waves is shown in Table B-1.

Table B-1: Expected Number of Completes at Each Wave

Wave	Completes		Total
	Main Sample	Oversample	
Wave 1	2,148	376	2,524
Wave 2	2,124 (with Chicago)*	330	2,454
Wave 3	1,912	297	2,209
Wave 4	1,720	267	1,987
Wave 5	1,548	241	1,789
Wave 6	1,394	217	1,611
Wave 7	1,254	195	1,449

*Note: Chicago joined the study in Wave 2.

Use of the internet and computer assisted telephone interviews increases our ability to

track and survey students across the seven-year period even when they are no longer in the school system. Sample maintenance strategies include: maintaining detailed contact information about the subjects and their families and two or three individuals who will likely have contact with them in the future; and sending birthday and/or holiday cards which will prompt notices of address changes if students move. When contact is lost, searches use internet resources such as the "Ultimates" (national white pages, email directories), and Google searches. In addition to standard tracking procedures, the research team uses current technology favored by youth such as text messages and monthly music downloads (which require students to provide a current email address to receive the download) to keep the students engaged in the study.

Estimation Procedure**.** For producing population-based estimates, each responding student is assigned a sampling weight. This weight combines a base sampling weight which is the inverse of the probability of selection of the student and an adjustment for nonresponse at the school level and the student level. The probability of selecting a student is the product of the probability of selecting the school district, the probability of selecting the school within the district and the probability of selecting the class in which the student is present. The inverse of the overall probability gives the base weight. Various selection probabilities are recorded and used to construct the sampling weight. The base weights are adjusted for nonresponse. All student level estimates including estimates of change are weighted estimates using the student weight. All student level analyses use student weights.

The objective is to select each student with a known probability of selection. Because of probability proportional to size (PPS) sampling at the first and second stages and unequal number of classes in selected schools, the overall probabilities of selection for students are

unequal. As indicated above, we determine the overall probability of selecting each student in the sample considering the three stages of sampling. The base sampling weight assigned to each student is the inverse of the overall probability of selection of that student.

The size measure for selecting primary sampling units using PPS sampling is total enrollment. The size measure for selecting schools offering grade 10 was enrollment in grade 10. We used PPS systematic sampling to select primary sampling units and schools within selected primary sampling units. The determination of probability of selection at each stage is straightforward under PPS systematic sampling. For example, the probability of selecting a PSU (say PSU j) within a Census division is

$$\pi_j = n\frac{x_j}{x}$$

where n is the number of PSUs selected, x_j is the total enrollment in PSU j and x is the total enrollment in all the PSUs in that Census division. Similarly, we can determine the probability of selection within a selected PSU. Classes were selected within a selected school using equal probability systematic sampling. As indicated earlier, the overall probability is determined by taking the product of the probabilities of selection at the three stages.

The adjustment for nonresponse at each stage is being done using the original base weights assigned to each unit. For example, the adjustment for nonresponse at school level involves the adjustment of school weights of responding schools such that the sum of the adjusted weights equal the sum of the weights of all selected schools including respondents and

nonrespondents. Similarly, the weights of the responding students are adjusted to account for nonresponding students. There is a final post-stratification adjustment of all student weights using a raking procedure such that the sum of the students in gender and race groups add to known number of students in the population of students in grade 10.

Thus, for producing population-based estimates, each responding participant is assigned a sampling weight. This weight combines a base sampling weight which is the inverse of the probability of selection of the participant and an adjustment for nonresponse at the school level and the student level. The probability of selecting a participant is the product of the probability of selecting the school district, the probability of selecting the school within the district and the probability of selecting the class in which the student is present. The inverse of the overall probability gives the base weight. Various selection probabilities were recorded and used to construct the sampling weight. The base weights are adjusted for nonresponse. All participant level estimates including estimates of change are weighted estimates using the participant weight. All participant level analyses also use participant weights.

Sampling Overview for NEXT Plus Substudy (N=560)

The sampling frame for the NEXT Plus substudy was all schools successfully recruited to participate in the basic survey. The following sampling stages were implemented.

1. In each of the nine strata (Census Divisions) all schools recruited were listed.
2. Geographic cluster sampling was used to group schools, which were in relatively close geographic proximity, into clusters (or “communities”).
3. On average, two clusters per Census Division were randomly selected for a total of 20 communities.

4. Within each “community” cluster, schools were first sorted by whether they were urban, suburban, and rural schools to assure representation.
5. Two schools within each cluster were then systematically sampled.
6. Each school selected contributed two classrooms that were randomly selected to participate in the basic survey.
7. At the study office, students’ in the selected classrooms were categorized as “overweight” or “normal weight” based on their height and weight measurements collected during the main study.
8. Seven overweight children and seven normal weight children were randomly selected across classes per school from the respective weight status categories and recruited to the substudy.
9. In Waves 4 through 7, participants turning 18 are re-consented with young adult consent forms.

Power Analyses

Basic NEXT Sample

The NEXT sample has adequate power to provide populations estimates with a margin of error of plus or minus 3 percentage points at the 95% confidence level. In addition, this sample enables sub-group analyses comparing Hispanic, African-American, and Caucasian youth. The oversample of minorities results in a final Basic Survey sample with a minimum of 200 Hispanic and 200 African-American participants. As indicated in the power analysis for the NEXT Plus subsample (below), this sample will enable sophisticated longitudinal comparisons across racial/ethnic groups.

NEXT Plus Sample

For specific hypotheses, the NEXT Plus subsample will be adequate to address primary hypotheses relating to obesity and cardiovascular disease. Power analysis and sample size estimation for specific hypotheses were conducted using Monte Carlo simulation procedures recommended by Muthen and Muthen (Muthen & Muthen, 2001). Monte Carlo simulation is the most common and preferred method to determine sample size for sufficient statistical power in multivariate analysis and structural equation modeling. In a Monte Carlo simulation, random samples with a specified sample size are generated repeatedly from a population with known parameters consistent with the proposed model. Path coefficients are then estimated from each simulated sample. The percentage of simulated samples that have significant parameters indicates the power of the study. The required sample size can be accurately determined by varying sample sizes in a series of simulations. The Monte Carlo study for determining power and sample sizes for the present study was conducted using Mplus version 3.0, which provides extensive simulation facilities for structural equation modeling.

The power analysis for determining sample sizes was conducted using a latent growth curve model for the relationship between participant physical activity and participant-reported peer physical activity, i.e., a linear model with seven repeated measures of physical activity as outcome with one-year intervals between the measures. Peer behavior was specified as a covariate with two additional covariates (gender and SES). Simulation was conducted using two peer effect sizes including various corresponding peer behaviors and outcomes in the study (substance use, physical activity, diet, obesity). A smaller effect size was defined by Cohen (1988) as 0.1 in standardized estimate and a medium effects size was 0.3. The path loadings from the intercept to the seven outcome measures were set at 1 and to the slopes were set from 0

to 7 with each unit represents a one year interval of assessment. Missing values were also generated in the simulation with each variable having 15% random missing.

Muthen and Muthen (2001) recommend several criteria for estimating appropriate sample sizes in power analysis for structural equation modeling. Parameter bias should not exceed 10%; standard error bias should not exceed 5%, and the coverage remains between 90 to 98%. The Monte Carlo simulation for this study conducted 1,000 replications with various sample sizes. The results from the simulation indicated that a final sample size of N = 440 for the linear model with small effect size had a statistical power of 96% to detect a peer effect, provided that missing values are random and below 15%. A separate simulation with medium effect size indicated that a sample size of N = 150 would have a power greater than 90% for detecting a peer effect. As a marker of clinical significance, a 0.3 to 0.5 SD between-group difference in physical activity should have a significant relation to health outcomes such as metabolic syndrome or adiposity. Thus, we would have the power to detect a clinically significant change in adiposity in analyses of the main sample and in analyses of selected subgroups. Subject retention has been higher in the NEXT Plus sample than the NEXT sample. The larger NEXT sample provides power to examine smaller effects within multilevel models and comparisons across sub-groups of interest. All criteria recommended by Muthen and Muthen (2001) were satisfied for the simulation studies.

B.2. <u>Procedures for the Collection of Information</u>

Waves 1 through 3 of data collection was conducted between January and July of 2010, 2011 and 2012, respectively . Waves 4 through 7 will be conducted during the same months of

2013, 2014, 2015 and 2016. Participants turning 18 after Wave 3 are re-consented with young adult consent forms.

Annual Surveys**.** The annual survey can be completed in approximately 60 minutes. The longitudinal survey focuses on a limited set of health behavior outcomes and has an expanded focus on potential etiological factors. Items are included when they are deemed essential to the outcomes of interest. Questions are drawn from items used in previous U.S. HBSC surveys or that have appeared in previous U.S. surveys, have evidence of good reliability and validity, and address unique issues related to the health of students in the U.S. Topic areas covered in the survey include:

- Eating habits, weight control, and body image;
- Physical activity;
- Sedentary Behavior and sleep;
- Substance use;
- Dating violence;
- Motor vehicle risks;
- Dental health;
- Family structure, environment, and communication;
- Peer influences;
- Medicine use and health care;
- Health status;
- Demographics.

Annual Survey Data Collection**.**

Participants turning 18 after Wave 3 are re-consented with young adult consent forms. Participants complete online surveys annually. To accomplish this, an email is sent to each participant with a secured, designated link to the online survey. Participants without access to a computer are given the opportunity to complete hard copies of the survey. Our studies show that there were no significant differences on the surveys due to response mode (HBSC forum; Seville, Spain, 2008).

Anthropometric Assessments (Waves 4 through 7): Assessments take place at participants' homes.

Online Dietary Recalls of the In-Home Sample (Waves 4 through 7). Although the NEXT surveys have a number of questions about diet, including eating at fast food restaurants and a brief food-frequency assessment for consumption of a few healthful and unhealthful foods, limits in the length of the survey do not permit estimates of daily caloric intake, proportion of calories from fat, carbohydrates, and protein, or whether daily intake meets dietary guidelines. To obtain these estimates, the NEXT Plus sample provides an additional dietary assessment each year. In Waves 4 through 7, the NEXT Plus sample complete the NCI ASA24, an online 24-hour dietary recall, for three days (random selection of two weekdays and one weekend day) each year. This method is completely consistent with NCI's recommendations for use of the ASA24 dietary recall. The ASA24 was developed by NCI to be consistent with the methods used in NHANES in-person 24-hour dietary interviews conducted by trained dieticians. More details on ASA24 can be found at: http://riskfactor.cancer.gov/tools/instruments/asa24.html and a demonstration of the instrument can be found at: https://asa24.westat.com/). The ASA24 was

developed by NCI and has been shown to have good reliability and validity for assessment of all nutrient groups.

Assessment of Physical Activity, Sedentary Behavior, and Sleep of the NEXT Plus Sample (Waves 4 through 7).

As is the case with dietary intake, because of within-individual variability of physical activity within a single day and across days, a single time sample may be inadequate to estimate individual levels of physical activity (Trost et al., 2000) and this variability may increase with age (Wickel et al., 2007). Although a single day may not be representative of a child's level of activity, there can be patterns across days. For example, there may be individual tendencies for higher levels of physical activity at particular times of day or days of the week (Trost et al. 2000). Thus, a week-long period is likely to capture this variability. For these reasons it is important to assess physical activity at different times of day and across multiple weekdays and weekend days. The number of days, the length of observation within each day, and the time of day sampled necessary to obtain a reliable estimate depends on the method of assessment as well as the age of the children being assessed. The recommendations for accurate and generalizable assessments are for up to 10 to 12 hours of observation per day, for minimums of three to 15 days depending on the assessment method, the level of physical activity necessary to meet the criteria for a particular intensity, and the age of the youth (Baranowski et al., 2008; Sirard and Pate, 2001; Trost et al., 2000). When physical activity is assessed with accelerometers, recommendations are for five to nine days of monitoring (Baranowski et al., 2008; Trost et al., 2000). We assess physical activity using an accelerometer for seven consecutive days. Patterns of weekend activity can also vary across ages; thus, sampling weekend days is important for

estimates.

Using multiple methods to assess physical activity increases the reliability, validity, and sensitivity of estimates of longitudinal changes in physical activity. Although the self-report items used in the NEXT survey have been shown to have good reliability and validity, self-report errors may be subject to systematic variation based on cognitive development; therefore, the errors associated with self-report may introduce an age bias in longitudinal changes based on self-report. In addition, physical activity can have significant daily variation. To address these potential problems, in Waves 1 through 7 physical activity is assessed over a 7-day period with an accelerometer with multi-day memory. During the home visit with the NEXT Plus cohort, the health researcher explains the use of the accelerometer and provides the participant with an accelerometer, written instructions, and a paid return envelope. Participants are also provided with a telephone number and email address in order to provide answers to any questions that arise during the week-long assessment.

One limitation of the accelerometer is that it cannot be worn during some sports activities, in the water, or when the participant is sleeping. The ActiWatch does not have these limitations (although watches may not be permitted during competition in some sports). The ActiWatch can be worn the entire day without concern for it getting wet and it provides minimal discomfort during sleep. During the home visit, each participant is also provided with an ActiWatch along with instructions on how to use it and how to return in the paid return envelope. Because the ActiWatch is worn on the wrist (the accelerometer is worn on the hip), it can over-estimate the energy expenditure of activities that primarily involve arm, rather than

trunk, movement. The primary reason for providing the ActiWatch is to obtain data on sleep; recent research suggests that adolescent sleep patterns affect obesity and mental health.

The accelerometer and the ActiWatch provide data on the frequency, duration, and intensity of bouts of physical activity. However, they do not provide information about the type of physical activity. An activity diary complements the activity monitor. For example, the diary tells us the precise activity that is reflected in the readings of the activity monitor, e.g., whether vigorous physical activity was due to participation in a sport (basketball), a leisure activity (jogging), or active transport (biking to the store). The diary differentiates going to bed, while the activity monitor indicates going to sleep. The diary also indicates the type of sedentary behavior (e.g., homework versus a video game). The diary provides context for specific behaviors (location, involvement of others) while the activity monitor provides a more precise measure of time of day, duration, and intensity. Together, they provide a much richer set of data on daily activity of the adolescents. Of course, these data can also be used for comparison of methodologies and contrast dimensions such as frequency, duration and intensity of physical activity when measured by self-report versus objective measurement. Each adolescent in the NEXT Plus sample is provided with a Physical Activity Recall form (see Attachment 7) and instructions on how to use it. The form has a date and time grid corresponding to the seven days when the accelerometer and ActiWatch are being worn. In addition to indicating dates and times the devices are worn, the adolescent indicates the primary activity within each grid including low energy expenditure activities such as sleeping, watching television, playing computer games, using the internet, and text messaging. A list of standard activities is provided. Data from the diary, the accelerometer, and the ActiWatch are linked to provide insight into activity

expenditure and the corresponding type of activities for the entire observation period.

NEXT Plus Assessments of Adiposity, Cardiovascular Risk, and Metabolic Syndrome (Waves 1 through 7). Home visits for collecting blood, blood pressure and other cardiovascular risk indicators are conducted in Waves 1, 4 and 7. Efforts at primary prevention of cardiovascular disease recognize the importance of serum cholesterol levels. Links between serum lipids and behaviors such as diet and exercise in adolescents deserve further research. Home visits are conducted at a time and place to accommodate the preferences of the participants.

Following standard protocols, fasting serum samples are obtained with a finger stick technique from the NEXT Plus cohort and collected in microtainer devices. 250 uL of serum is sufficient for the quantification of the lipid fractions and other assays. The biological markers obtained for obesity, cardiovascular disease risk, and metabolic syndrome include: fasting blood glucose, HbA1c, total cholesterol, triglycerides, LDL-C, HDL, C-reactive protein, uric acid, cotinine, height, weight, waist circumference, and blood pressure. Assessment of height, weight, and waist circumference follow the same protocols in each wave of the study. Blood pressure is assessed with a portable automated system. Blood samples are packed in ice and shipped to a central lab for analysis.

The only results that are immediately available are height, weight, waist circumference, and blood pressure. When blood pressures are in the at-risk range, youth (ages 18 and older) are told of these results along with a recommendation that they see their physician for subsequent evaluation and follow-up. When youth have high risk blood pressure values, youth are told to seek urgent care. When participants have at-risk and high risk levels of lipids and/or fasting

blood glucose, participants ages 18 and older are contacted with similar recommendations for seeking additional care.

B3. Methods to Maximize Response Rate and Deal with Non-Response

At initial consent and the conclusion of first-year assessments families/participants provided contact information for individuals who would know how to get in touch with them should they move. On a regular basis, cards are sent to families in order to remind them about upcoming assessments. These contacts reinforce the importance of their participation, thank them for their role in the study, and prompt them to work with the project coordinators when called. If mailings are returned by the postal service, the use of cards has the additional benefit of serving to notify investigators that families have moved. This allows the investigators to begin the tracking process before entering a critical period for subsequent assessments. When contact with a participant is lost for reasons other than withdrawal from the study, the tracking process begins with the individuals named by the families to provide contact information. Location efforts continue using internet resources such as the "Ultimates" (national white pages, email directories), and Google searches. In addition to these standard tracking procedures, the research team uses current technology favored by youth such as text messages and monthly music downloads (which require participants to provide a current email address to receive the download) to keep the participants engaged in the study. Participants also are asked to log on to the study website and provide updated contact information quarterly.

Main Study Incentives

The overall incentive structure for the annual survey is presented below:

Year of Participation	Completing Survey

Wave 4 – Post High School	$40
Wave 5	$50
Wave 6	$50
Wave 7	$50
Overall Total	$190

These incentives are consistent with other studies of adults and the time necessary to maintain contact with the NEXT research team and to complete the survey.

NEXT Plus substudy students

Year of Participation	**Completing home visit (height, weight, blood pressure, waist circumference, blood draw) and home surveys**	**Completing dietary questionnaire for three days**	**Wearing accelerometer and ActiWatch® for all seven days and complete activity diary**	**Total by Year**
Wave 4 – Post High School	$50	$50	$50	$150
Wave 5	No visit	$20/day	$60	$120
Wave 6	No visit	$20/day	$60	$120
Wave 7	$60	$20/day	$60	$180
Overall Total	$110	$230 for 16 days	$230	$570

School Response Rates and Nonresponse Bias Analysis

We contacted a probability sample of 137 schools and 80 (58.4%) agreed to participate in the survey. We conducted response bias analysis to determine if participating schools differed from schools that refused. The only significant difference between participating and non-participating schools was in the proportion of Asian American students. Because of the relatively small difference in the proportion of Asian American students in both groups (approximately 3%), this difference could have been due to the population of a single school in the refusal group

and/or the oversample of schools with a high proportion of African American students. The focus of recruiting was on obtaining sufficient subgroups of White, African American and Hispanic youth to permit comparisons; therefore, the differential effect of having a disproportional sample of Asian American youth should not affect planned analyses.

Participant Response Rates and Nonresponse Bias Analysis

Of 3,796 students recruited to participate, youth assent and parental consent were obtained from 2,619 (69.0%) students. Of those consented, 2,524 (96.4%) completed the Wave 1 survey. In Wave 1, confidential self-report surveys were administered by trained research assistants in the 10th-grade classrooms.

The retention rate from Wave 1 to Wave 2 was 87.1%. A comparison of Wave 2 dropouts with those retained in Wave 2 revealed no significant differences on race/ethnicity, gender, or the primary outcomes of interest (e.g., body mass index, physical activity, sedentary behavior, substance use, dating violence).

Due to a delay in obtaining final approval from the Chicago Public Schools, 246 students from Chicago Public Schools did not participate in the study until Wave 2. As a result, the Wave 2 sample was 2,454.

Nonresponse Bias Analysis in NEXT

Bias in a survey estimate because of nonresponse consists of two components. The first is the nonresponse rate and the second is the difference between respondents and nonrespondents in the population parameter that is being estimated. For example, if we are estimating a population percentage by selecting a simple random sample and computing the sample percentage and there is nonresponse, the bias in the sample percentage due to nonresponse is

given by

$$B(p) = (1 - r)(P_r - P_{nr})$$

where p is the sample percentage based on respondents, r is the response rate, P_r is the population percentage among the respondents and P_{nr} is the population percentage among the nonrespondents. Therefore, it is important to examine both the response rate and the differences between the responding and nonresponding groups in the analysis of bias in the estimates due to nonresponse. We describe below the steps that we followed for nonresponse bias analysis due to nonresponse by some schools in the NEXT sample. These steps are in accordance with the statistical standards set up by the National Center for Education Statistics (NCES) for nonresponse bias analysis (http://nces.ed.gov/StatProg/2002/std4_4.asp).

1. Examination of Response Rates

We examined both the overall response rate and the response rates for various subgroups as per the guideline 4-4-2A under NCES Statistical Standards. We examined school response rates by: (1) census division; (2) rural and urban; (3) enrollment (large schools vs. small schools); (4) proportion of minority students; (5) poverty index for schools; and (6) school type - public, Catholic and private schools. As indicated above, the only significant difference between participating schools and those schools that declined was for the proportion of Asian-American students (6% in non-participating schools; 3% in participating schools; $p < .05$). We have made appropriate weighting adjustments to reduce this bias.

We also examined the proportion of missing data among participants. The overall missing rate (e.g., due to a preference not to answer a particular question, a response of 'Don't Know', insufficient time or interest to answer the question, differences in reading skills) in Wave 1 was 9.7%. There were significant gender differences in missingness (males 10.2%, females 9.3%, $p < .001$) and significant differences by race/ethnicity with African-American and Hispanic youth having a higher rate of missing data (White 8.9%, African-American 10.5%, Hispanic 10.1%, Other 9.7%, $p < .001$). During Wave 2, surveys were completed outside of the classroom and would have had fewer time constraints. The overall missingness rate was lower in Wave 2, 8.4%. However, the gender and race differences persisted for gender (males 8.8%, females 8.0%, $p < .01$) and race/ethnicity but with African-American youth having more missing items than the other three subsamples (White 7.6%, African-American 9.6%, Hispanic 8.4%, Other 8.1%, $p < .001$).

2. Comparison of Sample and Frame Estimates

Per the NCES guideline 4-4-2C, we use sampling weight based on the probability of selection of responding schools without any nonresponse adjustment and data from the responding schools to compute population estimates of some characteristics available (not used for stratification at the time of selection of schools) on the sampling frame. These estimates are compared with the population values. If there had been large differences taking into account the sampling error, then this may have indicated bias because of nonresponse. We also generated estimates of students in responding schools by race/ethnicity, and compared this to the total computed from the population of schools on the frame to determine whether there was any bias in the estimates. This was not the case.

3. Comparison of estimates based on respondents to estimates from external sources

Per the NCES guideline 4-4-2C, we compared estimates of the prevalence of selected identical survey health behaviors items from the 2009-2010 Health Behavior in School-Age Children Survey of 10-grade students to determine whether there were large differences in the survey estimates. A large difference which cannot be attributed to sampling error might indicate a bias in the estimates. Although comparisons were only made when the survey items were identical in both surveys, this approach is limited as differences may not be solely due to sample bias.

The primary outcomes of interest in NEXT are behaviors related to obesity; these include physical activity, sedentary behavior, and diet. Responses to the Wave 1 NEXT survey for physical activity, sedentary behavior and diet did not differ significantly from responses to identical items on the HBSC survey. However, comparisons of substance use behaviors (there are no equivalent national surveys of dating violence or young drivers available for comparisons) indicated that the NEXT cohort reported a lower prevalence of smoking and alcohol use as well as lower reported use of Baltok, a fictitious 'drug' used to test dissembling. The fact that samples did not differ on physical activity, sedentary behavior or diet would suggest that there is little bias in the NEXT sample. Explanations for differences in reported substance use include: 1) the NEXT sample is indeed different from some national samples; and 2) the HBSC survey is anonymous while the NEXT survey is confidential but not anonymous – youth may have been more willing to report substance use in the HBSC survey, including fictional drug use.

The failure to find differences on key obesogenic behaviors and the likelihood that lower

reported substance use in the same cohort may have been due to the lack of anonymity suggests that there is little or no bias in the NEXT sample. Furthermore, because subsequent NEXT surveys are completed during the same time of year and there is no evidence that the concerns about anonymity will differentially affect subsequent responses, the cohort should be more than adequate for addressing the primary questions about the development of obesogenic behaviors, dating violence, substance use and driving.

4. Comparisons of Respondents by Successive Levels of Recruitment Effort

As per the guideline 4-4-2D by NCES, we compared schools that agreed to participate in the survey after the first contacts with those that agreed after several attempts or those that refuse first and then later agree. Estimates of student level characteristics were computed based on each successive wave of participating schools (i.e., adding respondents in the order of level of effort used to recruit the school) and the sampling weights based on probabilities of selection. If the estimates based on the initial sample and successively larger samples have a trend of either increasing or decreasing, this would indicate bias because of nonresponse. For example, if the percentage of students who are obese increased significantly as the number of responding schools increased, this might indicate that we are underestimating the percent of students who are obese. These analyses revealed no significant differences on the primary outcomes (e.g., body mass index, substance use) between students in schools that agreed to participate in 1 or 2 contacts versus 3 to 5 contacts versus >5 contacts before agreeing to participate in the study.

B4. <u>Tests of Procedures or Methods to be Undertaken</u>

This is a continuation of an OMB-approved study. All of the proposed modifications have been pilot tested, including the items in Peer Survey, which is a subsample of items used in the annual survey.

B5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

The role of outside consultants collecting the data, performing preliminary analyses, and staffing of this project is discussed in earlier sections of this application (B2, B3).

In addition to the Prevention Research Branch, our Division includes the Epidemiology and the Biostatistics and Bioinformatics Branches. All proposals and continuations of funding undergo extensive methodological review within the Division before moving forward.

Other consultants for this study include the research members of 40 HBSC countries who reviewed and recommended questions from the 2009/2010 HBSC study (OMB No.: 0925-0557, exp. date: 1/31/2012) according to their specialty interests, as members of HBSC focus groups. A core group of NEXT survey items is based on the HBSC survey. The HBSC Scientific Development Group required that all of the HBSC questions be piloted and reviewed externally before the questions could be included in the HBSC protocol. Besides the review of focus group questions, global external review was required under the HBSC protocol for significance of research topics, concepts, clarity of language used, and validity of measures to address those topics. Many of these reviews were completed by e-mail.

Consultations for this research project have been obtained incrementally since its inception. The initial concept and subsequent proposal were reviewed by two different External expert panels who evaluated the justification, design, and methods of the study. A third of the

panel included research methodologists or research statisticians. NICHD obtained external statistical review of five proposals for both methods and sample designs. The protocol, methods and assessments has also received multiple reviews by the NICHD Director of Intramural Research and panels of independent extramural investigators selected by the Director. Several levels of review and evaluation have been completed by participating institutes (NHBLI, NIAAA, and NIDA) including reviews by experts both internal and external to the National Institutes of Health. For example, in the review by the NHLBI Board of External Experts, the approval was near unanimous (with the one dissenting voter requesting the assessment of carotid intima-media thickness which has subsequently been added to the protocol). The proposal was also reviewed by the NHLBI advisory council.

As part of the IRB process, the proposal receives additional external reviews at NICHD organized by the Office of Intramural Research. Reviewers are drawn from three categories: longitudinal methodology; pediatrics; and pediatric cardiology. Reviews were glowing and without any criticisms of the methodology.

All changes in the project receive internal review by the Division of Epidemiology, Statistics, and Prevention Research and by the NICHD Office of the Director. The project is being reviewed again by the NICHD Board of Scientific Counselors in October, 2012.

Finally, the project receives an annual review by the NICHD Institutional Review Board.

In addition, all assessment procedures were distributed for review, comment, and endorsement to representatives of the broader education and health promotion community at the national, state, and local education agencies and those involved in the health and welfare of

children. These consultations included 31 representatives of state, local, and national education agencies.

The protocols, incentives, and surveys have been approved by the NICHD Institutional Review Board (IRB).