OMB Passback questions to NICHD on "Next" Generation Health Study
December 12, 2009

Overall Sample Design, Response Rates and Possible Biases

1. Please provide a description of the sampling frame from which the study is derived as well as what is known about its quality. We understand that NICHD chose not to use the commonly used Common Core of Data and Privacy School Survey at the National Center for Education Statistics (NCES), so would like to understand that quality of what we used instead.

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). Both for the 2005-2006 and 2009-2010 cross-sectional survey, the list of school districts and schools provided by QED was used. 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 require 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 go with the lists supplied by the QED for NEXT.

2. The researchers describe taking a sample of PSUs at the first stage (top of page 15), but elsewhere in the Supportive Statements, the PSU stage is presented as 14,000 districts, which is essentially the number of school districts in the United States. Please clarify.

The list of school districts was used to construct PSUs. PSUs were formed by grouping school districts within in each Census division. The total number of PSUs created for selecting public schools is **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.

3. What is the expected starting sample size of schools (i.e., to be approached about being in the study)? What is the expected school response rate? Table 1 on page 17 presents the estimated number of completes at each wave, but only refers to students without any estimation of school participation.

We will be selecting students for the longitudinal survey only in schools that agree to participate. We plan to contact a probability sample of 135 schools and expect that 90 schools will agree to participate in the survey resulting in a response rate of 67%. These estimates are based on our experience with recruiting schools for the HBSC survey.

4. What special procedures are in place to recruit private schools in sample, since they are not connected with the districts with whom NICHD been in contact, and historically have been hard to recruit?

Although we agree that private schools can present special recruitment challenges, we have had great success in recruiting them into the 2005-06 HBSC sample and the current 2009-10 HBSC sample. While the sample of private and parochial schools is proportionately smaller than the sample of public schools,

our recruitment rates for private schools are comparable to the recruitment rates of public schools. We use experienced recruiters and methods to explain the value of participating in the study and stress the need to make sure that "the voices" of private and parochial schools and students are adequately represented in the national findings. This strategy has been successful and we anticipate it will be no different for NEXT.

5. While this was discussed on the call, we wanted to go on record as indicating a concern about bias given that this study begins with 10th graders, rather than 9th graders. A significant percentage of students drop out of school before 10th grade.

We appreciate and to some extent share your concern. The main reason we are starting with 10th graders is to be able to follow them through the transition out of high school. Because the transition to college is a critical period for substantial changes in health behaviors and is understudied we made the difficult decision to start with 10th-grade students. Because of limitations on the duration of contracts, we were faced with either starting with 9th-grade students and not studying the transition out of high school or starting with 10th-grade students and including the transition. In addition, 9th grade is in middle school in some districts, which makes sampling and recruiting more difficult.

6. What is the expected starting sample size of 10th grade students (ie, those in classrooms in the study that will be asked to participate)? What is the response rate estimate for the first wave of data collection? Related, please describe of what universe the 80% and the 95% on page 16 refer. Please provide a justification for the expected response rate based on pretesting or similar studies.

The expected starting sample size of the 10th grade students is 3,580. This includes a sample of 3,150 students in the main sample and 430 in the oversample. If we contact 3,150 students in the main sample in participating schools, we expect that 80% of the students will agree to participate in the study. Of those that agree to participate we expect 95% to respond to the survey at the first wave. This gives an overall response rate of 76% yielding a sample of 2,394 respondents as shown in Table 1 of the supporting document. Similarly, we contact 430 students in schools that have a high proportion of minorities and 328 are expected to respond. The rate of 80% for retaining students contacted is based on previous studies which had a higher rate of retention.

This research team has already conducted successful nationally-representative, cross-sectional surveys of high-school students as well as a number of longitudinal studies of health behaviors of students. Use of the internet and computer assisted telephone interviews will increase our ability to track and survey students across the four-year period even when they are no longer in the school system. Other such longitudinal studies such as the NHLBI Growth and Health Study (NGHS) had a response rate of 78% but only focused on girls. NGHS's overall retention rate was 86% at year 5 and the same techniques for tracking students as well as new strategies (e.g., computer assisted tracking and follow-up with the schools) will be used to achieve a high retention rate (e.g., ~90%). Traditional strategies will include: follow-up with the schools; 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. If contact is lost, searches will begin using internet resources such as the "Ultimates" (national white pages, email directories), and Google searches. In addition to standard tracking procedures, the research team is exploring the use of social networking Web site such as Facebook or MySpace as well as current technology favored by youth such as text messages and monthly music downloads (which would require students to provide a current email address to receive the download) to keep the students engaged in the study.

7. Related, the table on page 7, SSB, does not seem to take into account eventual yield, or if so, is labeled unclearly. We need to understand the progression from universe to sample to respondents.

At each wave of data collection, we assume a retention rate of 80% and a response rate of 95%. The following table shows the progression from the initial sample of 3,150 students in the main to the sample of respondents. It is similar for the oversample. The numbers shown in the table are very close to the number of completes in Table 1

Expected Number of Completes at Each Wave

A. Wave	B. Number Selected	C. Agree to Participate (80% of B)	D. Respondents (95% of C)
Wave 1	3,150	2,520	2.394
Wave 2	2,394	1,915	1,819
Wave 3	1,819	1,455	1,382
Wave 4	1,382	1,105	1,050

8. In 75% of cases, two classes will be selected from each school. Why? What about the other 25%?

We have a sample of 90 schools. We want a sample of 3,150 students in the main sample. If we assume an average class size of 20 students, we need to select around 157 classes. If we select 2 classes in every selected school we will get 180 classes. Therefore we plan to select two classes from 67 schools (75%) and one class from the remaining 23 schools giving a sample of 157 schools. The resulting weighted sample will proportionately represent each region of the country.

9. Are the "1-2 classes are selected at random" within each school essentially a "home room" class? Will there be any checking to be sure that they are not "special" classes, such as all honors?

Principals are given the opportunity to select the "type" of classes to participate in the survey. They are instructed that the class type selected must be 1) a non-tracked class and 2) one that all 10th grade students are taking during that semester; the result is that all students should have an equal opportunity for being in a class that is selected to participate in the survey. For example, a principal may select to use health classes, PE classes, or homerooms as long as the above requirements are met. We use the same strategy in HBSC, and most principals chose to offer non-academic classes for random selection. Our experienced staff verifies that the requirements are met during their scheduling contacts with the school.

10. In school based studies that use multi-stage sampling and sample all students within a small number of classrooms, the typical precision analysis and overall data analysis plan take into account the nested, clustered nature of the sample, eg, via hierarchical linear regression modeling. Please describe how your study accounts for this design.

In this survey, it is of interest of obtain estimates of change in population percentages of health characteristics (yes/no) type between two time periods. The sample size should be large enough to be able to detect a difference of around 5 percentage points with 80% power when we do a two-sided statistical test. Assuming simple random sampling and a correlation of 0.5 between two time periods, it was estimated that a sample of 700 completes would be required

at the end of wave 4. The sample sizes are bigger than this number at earlier waves. Because the sample of students is selected using a multi-stage sampling design, the sample was increased to 1,050. This was done assuming an average design effect of 1.5. This value for the design effect was based on the 2005-2006 HBSC survey results.

11. Which students will be eligible to participate in the NEXT study? Students with various medical and cognitive conditions? Students with language challenges? What accommodations have been developed for such situations?

Students are excluded from the survey if they cannot read and understand the questions which are written in English or if they have developmental limitations that affect their ability to understand or provide age appropriate responses. If a student can cognitively understand the questions but cannot read or write responses, arrangements will be made to read to the students or assist in writing their responses, as we have in HBSC. Students with permanent physical disabilities that prohibit the collection of height, weight, or waist circumference will be excluded from the anthropometric measurements and the home substudy.

12. How many African-American students do the researchers expect will be recruited but not participate? At what stage of sampling are the researchers selecting additional samples of students? Differential response rates may be likely, but this possibility is not elaborated upon in the Supportive Statement. There is a passing reference on page 17 to selecting additional samples of students to screen and identify minority students.

It is anticipated that minority response rates for the NEXT Generation Health Study will mirror minority response rates seen in the HBSC Survey. The overall response rate to date in HBSC is 97.5%. The response rate in minority oversample schools (those with a high percentage of minority students) is 89.4%. Additionally, the percentage of students in minority oversample schools completing the survey is 85.5%, while 90.4% of students in the larger sample completed the survey. These rates indicate that although there is a differential rate in minority students, a high response and completion rate can be maintained in this population. The NEXT Generation Health Study will use the same student recruitment procedures employed in HBSC in order to replicate these response and completion rates.

The oversampling will be done at the time of data collection for the main sample in schools that have 30% or more of African American students enrolled. We estimate that we need to select an additional 430 persons to identify the African American students. This will be done in around 20 schools out of 90. The actual number of schools that we need for oversampling will depend on the number of African American students identified in the main sample. This will be monitored and on that basis the number of schools that we need to go to will be determined.

13. Page 21 suggests that when a participant has fewer than “3 completed assessments on a particular variable, that individual will be excluded from the longitudinal analyses involving that variable” yet the anthropometric assessments for most participants are only taken 3 times. Please clarify.

For some analyses, the 3 anthropometric assessments will be sufficient. In addition, we will gather self-reported heights and weights at four points in time and will be able to use the combination of self-report and measured heights and weights to impute more accurate estimates at the fourth point in time should it be necessary. In the subsample we will have measured height and weight at four points in time. These values will also contribute to the imputation equations.

The Biostatistics and Bioinformatics Branch (BBB) is within our Division and we work closely with branch investigators; for example, at least one investigator from the branch is assigned to every project within the division. Their statistical expertise will be employed to ensure that imputation of missing data is appropriately handled.

14. What appropriate methods will account for non-ignorable missing data? This problem will arise.

We agree that this problem is likely to need to be addressed. As noted above, we have available significant statistical expertise within our Division. Paul Albert, Ph.D, BBB branch chief, is an expert in the area of non-ignorable missing data and will identify the appropriate methods for dealing with these data. For example, depending on the nature of missingness, analyses could include methods such as shared random parameter models (Albert, 1999; Wu & Carrol, 1988) and selection models (Little, 1987).

Data Collection Procedures

1. Did we recall correctly that you say that 60-70% of the schools to be chosen have already been recruited? Which activities discussed under section B3 of the supporting statement have already been completed (e.g., the research applications to the special permission districts, the mailings to the superintendents prior to contacting the schools, mailing to districts, mailing to principals, telephone to principals.)

I was reporting the number of subjects already surveyed in the **HBSC cross-sectional sample** of students in grades 5 through 10. As of last week the proportion of students sampled is up to 76.8%. With regard to the NEXT Generation Health Study and section B3 of the SS, I checked with our contractor and they report that given the extended time required for review, research applications to special permission districts have been submitted. Informational mailings to selected States, districts, and schools have been provided or are scheduled with the understanding that actual recruitment into the study cannot be undertaken until approval has been received from OMB.

2. Page 12--Is the administrator (presumably school principal) the best source of information on health and physical activity programs. School administrators are notoriously busy and likely disinclined to take a survey which they deem the purview of another staff member (e.g., face validity) or not sufficiently important (e.g., limited 'free time', multiple competitors for attention). They may provide inaccurate information or not respond at all, creating bias if such non-response is systematic (e.g., principals in more challenging schools have less time to complete a survey).

The School Administrator Survey is discussed with the school principal. Principals are asked to identify the most appropriate person on their staff who can respond to questions about the school's health programs and curriculum. That individual is then contacted during data collection and asked to complete the survey. If the Principal determines that he/she is the best possible respondent the School Administrator Survey is completed by him/her. This same process has been executed in HBSC where the Administrator Survey response rate is currently 68%; however, data collection is still ongoing.

3. Total participation time is estimated at 40-45 minutes. If the main study survey takes 35 minutes, the sub-study an additional 5 minutes, then are there a mere 3-5 minutes devoted to reading instructions, explaining participant rights, and taking measurements?

In our pilot study with 9 14- and 15-year-olds, the average time to complete the survey was 35 min. including instructions. That will leave us approximately 10 min. to collect the anthropometric measures which are completed individually at a separate time. We are staffing the survey and anthropometric assessments with sufficient staff to efficiently move individual participants quickly through each stage.

4. The timeframe in which to collect the data from the approximately 38 students seems particularly tight. How were these numbers estimated? Are make-up days scheduled or planned, which, though necessary, will add time?

Please see above. The time to complete these measures is based on years of experience as well as pilot testing. Make-up days, when necessary, will not add to the burden of the participants but will add to the burden of the research staff. The potential burden on the school is one of the justifications for our incentive plan.

5. The SS takes a less than fully informed either/or approach to the survey mode of administration, ie, if it is school-based, then it must be paper-and-pencil. Several studies at NCES use computer administration for school-based surveys. Additionally, computer-based surveys are especially efficient when skip patterns are used.

There were several concerns that went into this decision. Because this is a longitudinal study, we felt it important to build rapport with the students. The same research staff that conducts the survey will also conduct the anthropometric measures. Secondly, we want to accommodate the schools as much as possible. Computer labs are heavily scheduled; schools are understandably reluctant to devote valuable resources to a health study. We always strive to accommodate the school schedule and needs when conducting in-school assessments. Finally, there is the cost of providing sufficient computer resources to conduct a national sample in a brief time frame; purchasing and coordinating the simultaneous delivery of sufficient computer resources to accommodate this project is not cost efficient. We do take advantage of computer technology in our other assessments and out-year assessments.

6. What mode affect studies will NIH conduct to be sure that response differences across waves are not the result of changing response mode, shown to be a significant factor for some types of sensitive topical areas?

We have examined this issue in a previous HBSC survey (2005/2006) and the HBSC consortium has actively addressed this issue; our studies show that there were no significant differences on the HBSC survey due to response mode (HBSC forum; Seville, Spain, 2008). We will continue to monitor this issue and address it again if necessary.

7. The SS indicates principal will introduce the survey to (presumably) selected parents. Is this an appropriate role for the principal? What is the expected response from principals to fulfill this plan? Then school principals will also be asked to help construct, distribute and collect the consent letters? Based on NCES's experience, principals are an extremely busy and will not help distribute letters for a study that is not related to state law.

School principals are asked to approve a school's participation in the study. We have modified our procedures based on our current experience with HBSC so that no longer will a principal provide a letter to parents introducing the study. The letter to parents will be from the NEXT Project Coordinator assigned to work with that particular school. The letter will indicate that the principal has approved the study but will not include any language indicating that the principal endorses or supports the study. At

no point is the principal asked to distribute or collect consent forms. The designated School Survey Liaison and classroom teacher are asked to assist with this task. Each of which is provided an incentive for their time and effort related to the distribution and collection of consent forms and related materials. This same process has been successfully executed in HBSC where the overall consent response rate is 97.5%.

8. Similar to the reluctance of school principals, teachers will not want to encourage students to return consent forms quickly. Likely, they will not do so. Is this why teachers are offered an incentive? No other reason is given in the Supportive Statements.

The designated School Survey Liaison and classroom teacher are asked to distribute and collect consent forms. This same process has been successfully executed in HBSC where the overall consent response rate is 97.5%. In recognition that this is an additional task for school staff and consistent with current and past OMB-approved procedures for HBSC, the School Survey Liaison is offered an incentive for the time and effort they contribute to the study. Our experience has been that staff members are very willing to take on this task. In cases where a district specifically prohibits using school staff to assist with the distribution and collection of consent materials, NEXT health **researchers** will visit the school and perform this task.

9. What is the alternative activity that the school administration will provide for students who do not participate? Is the alternative activity to keep the students in the same room so as to reduce distraction?

Prior to the school visit the assigned NEXT project coordinator will work with the designated School Survey Liaison to identify the most appropriate alternate activity for students in a randomly selected classroom who do not participate in the study. Examples of alternate activities used in HBSC include: students complete word games at their desk (provided by the research team) while other students complete the survey; students perform school work at their desk while other students complete the survey; students are relocated to a study hall in a separate classroom while other students complete the survey. These same options will be discussed with the School Survey Liaison at schools participating in the NEXT Generation Health Study.

10. SS B, Page 13--Keeping teachers in the classroom to monitor behavior may sound useful from a logistics standpoint, but unfortunately may make students wary about the confidentiality of their responses, thus eliciting less than truthful responses or higher non-response.

Asking classroom teachers to remain in the classroom during the survey administration to monitor student behavior is a strategy currently employed in HBSC. Project staff is trained to explain to the teacher that their assistance with monitoring student behavior is greatly appreciated but that in order to protect student privacy and confidentiality it is essential that he/she refrain from approaching students while they are taking the survey. Teachers are specifically asked to not move throughout the room while students are working on the survey or respond to student questions. The completed surveys are then placed in an envelope and sealed in front of the student to reinforce the confidentiality of their answers. This strategy is successful in HBSC and will be replicated in the NEXT Generation Health Study.

11. Please clarify where today NIH is in the list of school recruitment steps listed in SS B3. Using a calendar, and assuming a clearance date of 12/18, please clarify when each step has or will occur.

Special permission district applications have been submitted and are pending approval. State-, district-, and school-level informational mailings have been sent (or are scheduled), with the understanding that official recruitment into the study cannot be undertaken until after receiving OMB clearance. The following schedule outlines the anticipated final school and student recruitment activities, with the goal of first data collection dates beginning in mid-late February, 2010.

12/18/09 – OMB approval
Completed - Special Permission District Mailings
Completed - State-level Informational Mailings
Completed - District-level Informational Mailings
12/30/10 - Complete School-level Mailings
1/4/10 – Begin Recruitment Calls to Principals
1/11/10 – Begin Recruitment of Students at Endorsed Schools

12. Please also clarify how recruitment for "back up" schools fits into the schedule, once initial schools begin to indicate that they will not participate.

For every school that is randomly selected to participate in the study there is a reserve school in the same PSU that is available to serve as a replacement, if needed. Should a school elect to not participate the reserve school is immediately brought into the sample. The same notification and recruitment process (i.e., district mailings, school mailings, etc) is used for the newly added reserve school. This same process was used in HBSC and allowed for timely recruitment of replacement schools.

13. We have already discussed the need to revise and justify all proposed incentives. Both the size of the incentive to the schools and the size of the incentives to the teen participants are of concern from a cross-government policy standpoint. The amount of the incentive to the teen participant is also a concern from an ethical perspective. With respect to the latter, consider a joint parent/participant incentive for the in-home visit. Please make sure that each incentive is justified in terms of both burden and the likelihood of increasing participation. Any incentives designed for participant retention needs to be supported with the results of experiments. Also, pls reconsider teacher and administrator incentives, as they should be considered part and parcel of the 'school' incentive. Finally, please provide us with a table comparing the incentives being offered in this study and the ones being offered by the HBSC (including for the SHEPS).

We very much appreciated our previous discussion of incentives and now have a better understanding and appreciation of OMB's concern about keeping incentives reasonable and consistent across federal studies. Also, it seems from our conversation that we are not really far apart on what would constitute reasonable incentives for the NEXT Generation Health Study, given the substantial commitment involved. In the following paragraphs we provide brief justifications for the proposed incentives. We seek your guidance with respect to how best to move forward expeditiously to assure timely approval so that the study can get underway and without expensive delay.

In the past, OMB has been concerned about being able to recruit a representative sample of schools and, recognizing that it is not an educational survey, have actually encouraged us to increase our incentives to assure better recruitment and to recognize the burden being placed on schools resources. In the OMB-approved cross-sectional HBSC study, we are providing a school incentive of $500. Schools assist in distributing and collecting information about the study and consent forms. They provide updated information about classes and participants for sampling. They coordinate survey dates and times with us and arrange for appropriate in-school locations for survey administration. They complete a

school-level ('Administrator') survey. They accommodate our health survey procedures. And they assist with scheduling follow-up health surveys for absentees. The schools in the proposed study are being asked to do more than the HBSC study. In addition to the in-school health survey, they will accommodate two medical assessments: in coordination with our staff, they will provide space suitable for the anthropometric measures and collection of genetic material for all participating students; on a separate day, they will provide space for us to collect fasting blood samples and to provide breakfast to participants. In the following two years, schools will be asked to coordinate repeated anthropometric assessments of the participating students. Because the proposed study is longitudinal and more extensive than HBSC, the schools are being asked to make a greater, long-term commitment. Recognizing this additional burden and commitment to the study and the importance of recruiting and retaining a representative sample of schools, we are proposing an incentive of $1,000 in the first year and then return to the usual $500 incentives in the second and third years. The added demands placed on schools agreeing to participate in national health studies have been recognized in other government-funded longitudinal health studies. Among the best-known government-funded longitudinal studies of adolescent health, i.e. AddHealth beginning in 1992, provided an incentive of $1,000 to schools in the first year. A more recent government-funded study of adolescent health (HEALTHY, Drews et al., 2009) provided $2,000 to control schools in year 1, $4,000 in year 2 and $6,000 in year 3.

With regard to teacher and principal incentives ($50), our proposal is consistent with the OMB-approved HBSC surveys this year and previous years. In addition, our proposed incentives are consistent with the OMB-approved food and nutrition study conducted by the Food and Nutrition Service of the USDA. As with the proposed study, FNS/USDA conducted in-school anthropometric assessments. Our experience with HBSC is that some schools request that these incentives be provided as gift cards or as credits towards purchasing school supplies. We have always complied with these requests.

With regard to student incentives, we are mindful of the reviewers' concerns and have scaled down our individual incentives accordingly. Methodological studies of health assessments have indicated the importance of individual incentives for recruiting and maintaining a representative sample. With regard to initial recruitment in a health study, Martinson et al. (2000) found that a higher incentive ($15) resulted in significantly better recruitment than a low ($2) or no incentive. In the early 1990s, AddHealth provided $10 for completing surveys outside of school and $20 for each anthropometric assessment. Methodological analyses regarding maintaining samples of adolescents participating in longitudinal studies of health (Morrison et al., 1997; Drews et al., 2009) indicate the importance of incentives for maintaining a representative sample. They also suggest that lotteries might be effective for subject maintenance in longitudinal studies (accumulating greater chance of winning with increased number of responses). In light of this, we are proposing the following set of incentives.
First year: $0 for completing the in-school survey; $10 for completing each set of biological assessments (anthropometric, blood collection, blood pressure); $10 for wearing an activity monitor for a week; $10 for completing each dietary assessment; and $10 for completing a 7-day activity recall.
Subsequent years: $10 for completing online surveys; $10 for completing each set of biological assessments (anthropometric, blood collection, blood pressure); $10 for wearing an activity monitor for a week; $10 for completing each dietary assessment; and $10 for completing a 7-day activity recall.
Pending meeting federal and state guidelines, we propose a lottery for improving retention which rewards the adolescent for maintaining contact with the study. Of the students who update their information, ten students will be randomly selected each time to win a prize valued at $250.

14. Please also note that any use of a lottery requires written notification from HHS's Office of General Counsel that they are willing to work with NIH to ensure compliance with all applicable state laws and that they have confirmed that such compliance does not breach any confidentiality assurances.

Thank you. We were not aware of this requirement and will certainly comply.

15. The government cannot be seen as endorsing a particular product brand.

We will not mention specific products in any of our materials.

16. Page 11--Do parents also participate in the genetic marking tests?

No, only adolescent participants.

17. What types of methodological experts (as opposed to subject matter experts) participated in the expert review panels? Please provide minutes or other evidence that they provided useful input to the design of the study.

In addition to the Prevention Research Branch, our Division includes the Epidemiology and the Biostatistics and Bioinformatics Branches. All proposals undergo extensive methodological review within the Division before moving forward. The next review is a concept review by three external experts. This review is blind so we cannot indicate the participants but normally includes a methodologist and two content experts. The project is then cleared for a more extensive review by external experts as part of the contracting process. Unfortunately, I cannot reveal the names of the external reviewers because I have been advised by the organizers of these reviews that the names of the members of the review panel are confidential. However, I do know that about a third of the panel were research methodologists or research statisticians.

What follows is their review of the methodology:

"For Option 1 and Option 2 [this proposed study], they have both expertise and prior experience needed to carry on the project. They propose adequate sample size and use of probability sampling technique that will result in an acceptable national representative samples for the longitudinal purpose. One of the major key factors in conducting successful longitudinal studies is recruitment and retention of the subjects. The CDM/Abt group proposes specific plans and protocols, including specific incentives for student participation. The primary sample will be identified and screened so that an adequate number of major minority groups (Black and Hispanics) will be included in the study. The team spells out their detailed plan and steps needed for conducting the longitudinal options.

They propose to conduct single field training for the field workers for the survey. This can save some resources and contribute to cost effectiveness. They are very mindful of protection of human subjects and have experience and protocol for that.

The CDM/Apt proposal is reasonably strong and detailed with a plan of action with regard to cross-sectional and longitudinal procedures, survey scheduling, delivery, data collection, quality control, and providing a report.

These studies suggest previous experience in recruiting, retaining and tracking a large number of schools over time with high participation rates and timely survey completion. CDM is very experienced in obtaining school and school district agreement to participate in a timely fashion. These investigators have a documented track record of subject recruitment, maintenance, surveying, and confidentiality issues in longitudinal studies of adolescents and children.

CDM is forthcoming in their understanding of potential problems and solutions (page 9-10) to conducting this study in a timely fashion. In turn, CDM has a plan to overcome these obstacles based on their past experience. These investigators have a clear understanding of the project objectives and the appropriate methodology and personnel to conduct such research in a highly professional manner while protecting the rights of human subjects in research.

Weaknesses:
While a team approach of multiple agencies can be strength of the proposal since it brings together diverse expertise, its coordination can be problematic.

For a project of this magnitude, one should expect at least a few letters of endorsement or support from school affiliates or from local, state, or national agencies."

The next review was multiple reviews by internal and external review groups at NHLBI. Internal investigators included:

Lawrence Fine, M.D., Dr.P.H., Chief, Clinical Applications and Prevention Branch, Division of Cardiovascular Sciences, NHLBI
Michael Lauer, M.D., Director, Division of Cardiovascular Sciences, NHLBI
Denise Simons-Morton, M.D., Ph.D., Director, Division for the Application of Research Discoveries, NHLBI
Dr. Susan Shurin (acting director NHLBI) chaired one of the NHLBI internal reviews - the Idea Forum.

The NHLBI Board of External Experts (BEE) reviewed this proposal. The members of this board included:

Edward J. Benz, Jr., M.D.
President
Dana Farber Cancer Institute
Dana 1628
44 Binney Street
Boston, Massachusetts 02115
617–632–4266 ♦ 617–632–2161 (fax)
E-mail: edward_benz@dfci.harvard.edu

Eric Boerwinkle, Ph.D.
Director, Human Genetics Center
University of Texas Health Science Center at Houston
1200 Herman Pressler Drive, RAS E447
Houston, Texas 77030
713–500–9800 ♦ 713–500–9264 (fax)
E-mail: Eric.Boerwinkle@uth.tmc.edu

Eugene Braunwald, M.D., M.A.C.C.
Chairman
Thrombolysis in Myocardial Infarction Study Group
Brigham and Women's Hospital
350 Longwood Avenue, First Floor
Boston, Massachusetts 02115
617–732–8989
E-mail: ebraunwald@partners.org

William W. Busse, M.D.
Professor of Medicine
Department of Allergy and Clinical Immunology
University of Wisconsin Hospital
K4/912 CSC 9988
600 Highland Avenue
Madison, Wisconsin 53792
608–263–6183 ♦ 608–263–3104 (fax)
E-mail: wwb@medicine.wisc.edu

Robert M. Califf, M.D.
Vice Chancellor for Clinical Research
Director, Duke Clinical Research Institute
Professor of Medicine
Duke University Medical Center
2400 Pratt Street, Room 0311, Terrace Level
Durham, North Carolina 27705
919–668–8820 ♦ 919–668–7103 (fax)

E-mail: calif001@mc.duke.edu
George M. Church, Ph.D.
Professor of Genetics, Harvard Medical School and Harvard-MIT Program in Health Sciences and Technology
HMS New Research Building, Room 238
77 Avenue Louis Pasteur
Boston, Massachusetts 02115
617–432–7562 ♦ 617–432–6513 (fax)
E-mail: gmc@harvard.edu

Barry S. Coller, M.D.
Professor of Medicine
David Rockefeller Professor
Rockefeller University
1320 York Avenue
New York, New York 10021
212–347–7490 ♦ 212–347–7493 (fax)
E-mail: collerb@rockefeller.edu

Allen W. Cowley, Jr., Ph.D.
Professor and Chair
Department of Physiology
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, Wisconsin 53226
414–456–8277 ♦ 414–456–6546 (fax)
E-mail: cowley@mcw.edu

James D. Crapo, M.D.
Professor
Department of Medicine
National Jewish Medical and Research Center
1400 Jackson Street
Denver, Colorado 80206
303–398–1436 ♦ 303–270–2243 (fax)
E-mail: crapoj@njc.org

Ronald G. Crystal, M.D.
Director and Professor
Institute of Genetic Medicine
Cornell University
Joan and Stanford I. Weill Medical College
1300 York Avenue, Box 96
New York, New York 10021
646–962-4363 ♦ 646–962–0220 (fax)
E-mail: rgcryst@mail.med.cornell.edu

Karina W. Davidson, Ph.D.
Herbert Irving Associate Professor
in Medicine and Psychiatry
Co-Director
Center for Behavioral Cardiovascular Health
622 West 168th Street, PH9 Room 948
Columbia University Medical Center
New York, New York 10032
212–241–3800 ♦ 212–241–4000 (fax)
E-mail: kd2124@columbia.edu

Serpil C. Erzurum, M.D.
Chair, Department of Pathobiology
Lerner Research Institute (NC22)
The Cleveland Clinic Foundation
9500 Euclid Avenue
Cleveland, Ohio 44195
216–445–6624 ♦ 216–445–6625 (fax)
E-mail: erzurus@ccf.org

Gary H. Gibbons, M.D.
Director, Cardiovascular Research Institute
Professor of Medicine
Morehouse School of Medicine
720 Westview Drive, South West
Atlanta, Georgia 30310
404–752–1545 ♦ 404–752–1042 (fax)
E-mail: ggibbons@msm.edu

David Ginsburg, M.D.
James V. Neel Distinguished University
Professor of Internal Medicine and
Human Genetics
Division of Molecular Medicine and
Genetics
Department of Internal Medicine
University of Michigan Medical School
210 Washtenaw Avenue, Room 5028
Ann Arbor, Michigan 48109–2216
734–647–4808 ♦ 734–936–2888 (fax)
E-mail: Ginsburg@umich.edu

Geoffrey S. Ginsburg, M.D., Ph.D.
Director, Center for Genomic Medicine
Institute for Genome Sciences and Policy
Professor of Medicine
Duke University Medical Center
101 Science Drive, Box 3382
Durham, North Carolina 27708
919–668–6202
E-mail: geoffrey.ginsburg@duke.edu

Christopher B. Granger, M.D.
Associate Professor
Division of Cardiovascular Medicine
Department of Medicine
Duke University Medical Center
101 Science Drive, DUMC 3409
Durham, North Carolina 27710
919–668–8900 ♦ 919–668–7056 (fax)
E-mail: grang001@mc.duke.edu

Philip Greenland, M.D.
Harry W. Dingman Professor
Executive Associate Dean for Clinical
and Translational Research
Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University
680 North Lakeshore Drive, Suite 1102
Chicago, Illinois 60611
312–908–7914 ♦ 312–908–9588 (fax)
E-mail: p-greenland@northwestern.edu

Judith S. Hochman, M.D.
Director, Cardiovascular Clinical Research
Harold Snyder Family Professor of
Cardiology
New York University School of Medicine
550 First Avenue, OBV-A634
New York, New York 10016
212–263–6927 ♦ 212–263–7129 (fax)
E-mail: Judith.Hochman@nyumc.org

Isaac S. Kohane, M.D., Ph.D.
Director
Countway Library and Center for
Biomedical Informatics
10 Shattuck Street, 5th Floor
Boston, Massachusetts 02115
617–919–2184

E-mail: isaac_kohane@harvard.edu

Deborah Nickerson, Ph.D.
Professor of Genome Sciences
Adjunct Professor of Bioengineering
University of Washington
Foege Building, S213A
P.O. Box 357730
1705 Northeast Pacific
Seattle, Washington 98195-5065
206–685–7387 ♦ 206–221–6498 (fax)
E-mail: debnick@u.washington.edu

Garry H. Nolan, Ph.D.
Associate Professor of Microbiology and Immunology
Baxter Laboratory in Genetic Pharmacology
Department of Microbiology and Immunology
Stanford University School of Medicine
Center for Clinical Sciences Research
269 Campus Drive, Room 4216
Stanford, California 94305-5175
650–725–7002 ♦ 650–723–2383 (fax)
E-mail: gnolan@stanford.edu

Daniel Rader, M.D.
Professor of Medicine, Pharmacology, and Pathology and Laboratory Medicine
University of Pennsylvania School of Medicine
654 BRB II/III 6160
421 Curie Boulevard
Philadelphia, Pennsylvania 19104
215–573–4176 ♦ 215–573–8606 (fax)
E-mail: Rader@Mail.Med.Upenn.Edu

Paul M. Ridker, M.D.
Director, Center for Cardiovascular Disease Prevention
Preventive Medicine Division
Brigham and Women's Hospital
900 Commonwealth Avenue
East Boston, Massachusetts 02215
617–732–8790 ♦ 617–734–1508 (fax)
E-mail: pridker@partners.org

David Scadden, M.D.
Professor of Medicine
Center for Regenerative Medicine and Technology
Massachusetts General Hospital
185 Cambridge Street, CPZN-4265A
Boston, Massachusetts 02116
617–726–5616 ♦ 617–724–2662 (fax)
E-mail: scadden.david@mgh.harvard.edu

Deepak Srivastava, M.D.
Director, Gladstone Institute of Cardiovascular Disease
Professor, Departments of Pediatrics and Biochemistry and Biophysics
Wilma and Adeline Pirag Distinguished Professor in Pediatric Developmental Cardiology
University of California, San Francisco
1650 Owens Street
San Francisco, California 94158
415–734–2716 ♦ 415–355–0141 (fax)
E-mail: dsrivastava@gladstone.ucsf.edu

James T. Willerson, M.D.
President
Texas Heart Institute
St. Luke's Episcopal Hospital
6770 Bertner Avenue
Houston, Texas 77030
713–500–3010 ♦ 713–500–3059 (fax)
E-mail: jwillerson@sleh.com

In addition, the project was reviewed by the NHLBI advisory council. Members of this council can be found at: file:///C:/Documents%20and%20Settings/iannottr/Local%20Settings/Temporary%20Internet%20Files/Content.Outlook/DWFS9NIZ/National%20Heart%20Lung%20and%20Blood%20Advisory%20Council%20%20Members.htm

Finally, as part of the IRB process, the proposal received another external review at NICHD organized by the Office of Intramural Research. Again, this review was blind so we cannot list reviewers but we do know they were drawn from three categories: longitudinal methodology; pediatrics; and pediatric cardiology. Two of the reviews were glowing and without any criticisms of the methodology. The third arrived after the OMB application was submitted and was more critical. Similar to those comments expressed in the OMB review, this reviewer requested additional details about sampling and recruitment; however, consistent with other reviews, this reviewer was very positive about our incentive plan:

"Importance - significance to science or clinical practice.
US participation in the international HBSC survey effort is critically important to an understanding of adolescent health behaviors and the social and behavioral factors that shape adolescent health in the US. The cross-national comparative value of US participation makes this a high priority project for funding. In addition, the range of data proposed to be collected will provide researchers with the ability to study associations among a set behavioral and health variables that have not been examined before. Particularly valuable is the longitudinal component of the data collection, which will provide opportunities to explore how behavior early in adolescence can shape behavior in early adulthood.

Adequacy of research design
There are several flaws in the presentation of the research protocol, particularly in the sample design.

It is difficult to replicate the sample size and power calculations, although the text covers standard components and assumptions needed for the calculations.

The oversampling rates and procedures on page 15 are not described in adequate detail.

The nature of the primary sampling units is unclear, and why a stage of selection at the school district level is being used in not justified.

The discussion of estimation procedures on the bottom of page 15 and top of page 16 suggests that selection is unequal probability. The three stages of selection would appear to yield unequal probabilities of selection for students, but without the probability specification this reviewer could not be sure that the selection would or would not be equal chance. The probability proportionate to size selection procedure is not adequately described (for example, the measures of size are not given for each stage - enrollments could be total number of students, or only those in 10th grade), nor is there a formulation of the probabilities of selection for students across the three stages. There is a reference to developing non response adjustment weights without sufficient specification.

The protocol for recruiting schools is also difficult to follow. There are indications that parental and student consent will be sought, but methods for securing parent and student assent are not described. The process of mailing notification to sample school districts and schools may yield lower response rates than anticipated, and personal visits should be included as part of the recruitment protocol. It is not clear

if school district and school level nonresponse has been accounted for in the sample size calculations. It is not uncommon in school surveys to substitute for nonresponding schools, or school districts, but the issues of school district and school level nonresponse is not discussed.

Fortunately, it appears that the project will be carried out with the assistance of a reputable consulting firm that is known to have experienced and capable survey sampling statisticians and field data collection capabilities. The flaws in the presentation are less of a concern given the abilities of the survey data collection firm in mind. Given the deficiencies in the presentation of the survey protocol, the investigators are urged to rely on the expertise of a reputable firm, like the one mentioned on page 17, to complete the design, carry out data collection, and generate a data set that has suitable weights and other features (sampling error codes for variance estimation, imputation for item missing data).

Another favorable feature of the proposed research is a well thought through set of incentives for school districts, schools, and students.

Additional Comments
The importance of this work makes it vital that this project be funded in one form or another. Despite the flaws in the proposal, the data collected, and the research questions to be addressed, are too important not to fund this project."

18. Please provide all letters and scripts that accompany the various data collection activities and questionnaires.

Scripts for each data collection activity have been incorporated into the relevant Attachments.

Study Subsample and methods

1. SS A, Page 11--If "a subsample of consenting students will participate in a more extensive assessment" – how are these students selected? On what basis? Does this introduce the potential for selection bias?

In SSB, section B1, page 10, the following description of the selection of the subsample of students is provided.

The sampling frame for the selection of the sample of schools for the substudy will be all schools successfully recruited to participate in the basic survey. In each of the nine strata (Census Divisions) all schools recruited will be listed. Schools, which are in relatively close geographic proximity, will be grouped into clusters (or "communities"). Clusters will be formed such that these will be approximately equal in size in terms of the number of students. On average, two clusters per Census Division will be randomly selected for a total of 27 communities. Within each "community" cluster, schools will be listed and sorted by whether they are urban, suburban, and rural schools to assure representation in the sample. Using systematic sampling, two schools will then be selected from each of the 27 communities to provide the sample of 54 schools. Students in the one or two classrooms that were originally randomly selected to participate in the basic survey in that school will be eligible for selection in the subsample. (Expected number of students ~19 students/class; ~38 students/schools with two classes). At the study office, students' in the selected classrooms will be categorized as "overweight" or "normal weight" based on their height and weight measurements collected during the main study. Seven

overweight children and seven normal weight children will be randomly selected across classes per school from the respective weight status categories and recruited to the substudy.

2. Additional concerns arise with the data collection method. Students will recall their diets on "three separate days in each year of the study." How will these days be selected, at random or systematically? How does this method vary from that used by NCI's dietary researchers? Is the recall instrument the same one developed by NCI for NANES?

Our method of selecting days (random selection of two weekdays and one weekend day) 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. The ASA24 has had extensive reliability and validity evaluations during its development.

3. Will the students' activity diary be used to confirm or to refute what the accelerometer and the ActiWatch record? Will this tap into the difference between adolescents' perceptions of their activity and a more objective recording of this exertion? Or will the adolescents believe that they are being asked to repeat the recording of the equipment and not bother to complete the diary, so one serves as the back-up to the other?

The activity diary will complement the activity monitor. For example, the diary will tell 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 will differentiate going to bed, while the activity monitor will indicate going to sleep. The diary will also indicate 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 objectively.

4. Given the effect of the burden on data quality (regardless of the size of the incentive) and the likelihood of diminishing returns in terms of variability for school students after three days of activity monitoring, please explain why seven days of activity log and physical assessment is optimal.

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 assessment 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 for school-age children (Baranowski et al., 2008; Trost et al., 2000). We propose to assess physical activity using an accelerometer for seven consecutive days. Patterns of weekend activity can also vary across ages; thus, sampling weekend days may be important for estimates.

5. The bullet at the bottom on page 10 of SSA lists the biological measures to be investigated. Are you planning to store any of the blood for future analysis (e.g., if new genetic markers are identified?) If so, how do you address this on your consent forms?

Our consent form indicates that genetic data will be stored to measure these specific markers and provides parent and child with the ability to opt-out of having the child's genetic material collected. We will not conduct a GWAS with the genetic samples. Instead we will look for specific genetic markers including those for obesity, cardiovascular disease, metabolic syndrome and substance use. Because this is a rapidly developing field, we plan to store the genetic material until the end of the study to permit assessment of the latest significant markers.

6. With respect to genetic markers in general, will whole genome analysis be performed on these samples?

No. Please see above.

7. Page 17--Can NICHD justify the target numbers of students for each group in the substudy? Seven overweight children and seven normal weight children will be randomly selected across classes from 54 schools. How are the 54 schools selected from the 80? How are the students selected from the classrooms? What is the retention rate that you expect for this subsample?

<u>Justification for the substudy sample size</u>.

For specific hypotheses using data from the substudy, the subsample of the longitudinal sample 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, 2000). 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 student physical activity and peer physical activity, i.e., a linear model with four 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 four outcome measures were set at 1 and to the slopes were set from 0 to 4 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 = 550 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 subgroup 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 should be higher in the in-home assessment than the in-school sample because they will have already completed the Year 1 in-school assessment and will have consented to the additional in-home assessment. To assure a final sample size of 550 we will start with a sample of 750 in Year 1. The larger sample participating in the survey but not the home visits would provide 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

How are the 54 schools selected from the 80?

In each of the nine strata (Census Divisions) all schools recruited will be listed. Schools, which are in relatively close geographic proximity, will be grouped into clusters (or "communities"). Clusters will be formed such that these will be approximately equal in size in terms of the number of students. On average, two clusters per Census Division will be randomly selected for a total of 27 communities. Within each "community" cluster, recruited schools will be listed and sorted by whether they are urban, suburban, and rural schools to assure representation in the sample. Using systematic sampling, two schools from each of the 27 clusters will be selected to obtain the sample of 54 schools.

How are the students selected from the classrooms?

Students in the one-two classrooms that were originally randomly selected to participate in the basic survey in that school will be eligible for selection in the subsample. (Expected number of students ~19 students/class; ~38 students/school with two classrooms). At the study office, students' in the selected classrooms will be categorized as "overweight" or "normal weight" based on their height and weight measurements collected during the main study. Seven overweight children and seven normal weight children will be randomly selected across classes per school from the respective weight status categories and recruited to the substudy.

What is the retention rate that you expect for this subsample?

We anticipate a maximum 10% attrition rate at each wave with an end sample of 547 participants.

Topical Content and Questionnaires

1. Page 25, Research Question 6, what does the nature of peer influence mean? And how do the questionnaire items map onto this construct? It is not clear from either the questionnaire or the Supportive Statement.

I apologize for using such a vague term. We are measuring a number of dimensions of peer behavior and potential peer influence. However, the point of this hypothesis is to examine whether the relations of substance use with these different aspects of peer relationships and of peer behavior are different for different substances. For example, peer behavior may be very important for social drinking but may be less important for smoking. Or there may be a difference across substances as to whether participating in a behavior with peers is more or less influential relative to whether the peer engages in that behavior.

2. In Research Question 8, whose policies are affecting substance use? And where are policies captured?

Again, more precision may have helped. We have previously conducted similar research. We will be looking at school policies (captured in the administrator survey) as well as the policies of local communities which are available in existing databases (though variation within communities is more likely for regulations and enforcement of policies regarding tobacco than those for other drugs).

NEXT Generation Health Study Survey 2009-10 Main Questionnaire

1. The last part of the instructions which enumerates the exceptions for response patterns is unnecessary and confusing.

Thank you. We agree. Pilot testing of the survey has provided feedback that concurs with this statement. The survey instructions will be appropriately modified.

2. Question 2--For adolescents living in joint custody arrangements, for which house do they have a computer, games console, or television in the room where they sleep? Should a "usually" be inserted to specify?

Thank you for the suggestion. This question was developed by one of our collaborators and can be modified. We would like to insert 'usually' before 'sleep'.

3. Question 5 --"It is a choice I really want to make for myself." This response emerges several times throughout the questionnaire, but we wonder about its meaning and utility. First, it sounds artificial. Second, it is a needlessly complicated means of stating "I want to."

We agree. This is a question we adapted from the extensive research on Self-Determination Theory (SDT). We applied responses that previously have been validated in different areas of health behaviors (http://www.psych.rochester.edu/SDT/questionnaires.php). During our pilot test of the survey we conducted cognitive interviews with the respondents and they pointed out the same problem. We would like to replace this response and the same response in the parallel questions (24, 31, 62, 63, 70, 71, and 72) with "It's a decision I make on my own".

4. Further in Question 5, "I have the opportunity or it is part of how my day is structured." The first half sounds like the opportunity is accidental, purely fortuitous. The second half sounds like the activity is planned. Combining the two into one statement to reject or accept will confuse not only respondents but also analysts.

The response is not designed to capture which external control is prevalent but rather whether the respondent views the behavior as externally controlled. Respondents in our pilot study interpreted this response correctly and did not express any problems with it.

5. Why not include a response option of "my friends do it"?

We assess whether their friends do it later in the survey and will use those items to determine whether friends' health behavior indirectly influences the respondent's health behavior. This item will assess whether they believe that friends' approval affects their behavior.

6. Question 12 -- What does "somewhere else" mean in the final response option? Delete?

This is a standard HBSC item and has been used successfully in surveys of hundreds of thousands of adolescents. Within the US, students have answered that they live with an aunt or uncle, live in a juvenile detention center (but attend school outside the center), etc.

7. Question 24 – Again, "it is a choice I really want to make for myself" seems awkward, and in this case, largely indistinguishable from "It is personally important to me." Also, why is the response about planning the day for physical activity -- "My day is structured for it" -- missing here? That seems relevant and should be included.

We agree and would like to change the first response (see item 3 above). As for the last response, we are using a standard SDT response to assess external control.

8. What is the demonstrated reliability and validity of Question 28? Absent evidence to the contrary, this snoring question seems too subjective and too imprecise to capture a sleep apnea diagnosis, even crudely.

These are essentially identical to questions asked in NHANES. Data from Sleep Heart Health and Wisconsin Sleep Cohort studies demonstrate that these are valid and reliable questions (Hu et al., 1999; Sassani et al., 2004; Villaneuva et al., 2005; Young et al., 2008).

9. Question 31—Perhaps there should be an option to capture "I am required to do so." Why is the option for "My parents and family tell me to do it" missing? This seems crucial, especially since parents and friends may be the primary dispensers of food and dining companions.

See explanation provided for items 3 and 7.

10. Question 34--Is a breakfast consisting of oatmeal or cream of wheat a hot meal? Along with a slice of pizza for lunch and some meat and veggies for dinner, that equals to 3 hot meals a day. In that case, which is likely frequent, does a respondent darken 3 circles? Is this trying to get at a "main meal?"

This is a question that has previously been approved by OMB and has been used in HBSC surveys with hundreds of thousands of adolescents in the US and internationally. Because these questions have already been evaluated for reliability and validity and have appeared in numerous peer-reviewed professional journals, we are reluctant to make any changes which might threaten their validity or scientific value. However, the stem asks where they 'usually' eat a hot meal, so it should not pose a

problem for students that sometimes eat in a restaurant; eating multiple meals in a day would still be in two locations.

11. Questions 44 through 52 -- Is this battery of questions (or a subset of it) supposed to tap psychological well-being? Is it an established battery fur use with teenagers?

Yes, this is an established battery for use with adolescents. Questions 44 through 52, with the exception of question 49, have previously been approved by HBSC and/or OMB and have been used in HBSC surveys with hundreds of thousands of adolescents in the US and internationally. Because these questions have already been evaluated for reliability and validity and have appeared in numerous peer-reviewed professional journals, we are reluctant to make any changes which might threaten their validity or scientific value. Question 49 is from the CDC Center for Epidemiologic Studies Depression Scale (CES-D) (Dahlberg et al., 1998).

12. Questions 46, 47, 48 – most of the federal statistical system (NCHS, Census, BLS, BJS) have moved away from characterizing disability in terms of illness or condition and favoring questions that get at the functional limitations that the person faces. We can provide a copy of those questions. Of concern here in what the response to question 48 actually tells you – does 'affect attendance or participation' include attending a special school to which transportation is provided? Does it include limits to participation in physical activities, but having a wheel chair that makes it possible to attend all classes except physical education, etc.

Please see response to item 11 above. We are interested in overall health but also specific illnesses that may relate to obesity or heart disease or require medications that increase risk of these health problems rather than functional limitations but ask the question more broadly. We will not be recruiting students from special schools but do include students with physical disabilities. We would definitely appreciate receiving copies of the questions you have identified as well for potential inclusion in subsequent surveys.

13. Question 50 --What about grandparents, especially if that has been noted as the primary living arrangement? What about other adults at school?

We appreciate the suggestion. These are standard questions that have been used in OMB-approved surveys. The focus is on parental support, not general support from an adult.

14. Question 55 -- This should be rewritten into the form of a question.

Thank you for catching this. It must have been caused by a copying error and yet missed in review by numerous sets of eyes. We would like to replace the stem with: "Is your group of friends well accepted by your parents?"

15. Is there are reason to lump together 'Almost Never' and 'Never'?

This has been done with several questions because it is unusual for adolescents to say 'never' and makes this choice represent infrequent.

16. Questions 56 - 58 – Please justify the need to ask for up to 5 friends of each gender. Please also justify the choice of sub-questions, including the extent to which they do and do not overlap across the questions.

Our pilot adolescents were like-minded, indicating that they felt compelled to list 5 even though the question says 'up to five'. As a result, we will reduce these to 3 friends of each gender. The different sub-questions are derived from the AddHealth study and are needed to determine the extent of the friendship and the involvement of friends in various health behaviors and contexts for preventive and risky health behaviors.

17. Question 60 -- The last time you were in a physical fight…. What about an altercation with a classmate? What about fisticuffs with an acquaintance? These seem more likely than a fight with a total stranger.

'Total stranger' is but one of several choices which include 'someone I know', i.e., an acquaintance.

18. Question 62--The response options or items do not match the construct the question is intended to measure. "I threaten to hurt a romantic partner, because it is a choice I really want to make for myself?" That makes no sense. Who will endorse "It makes me feel good"? This is the one spot in which the item responses should be drastically overhauled and not necessarily match the other items with similar structure. The following likely responses should be included:
    - I want to impress my peers
    - I get mad at the person.
    - I want to make my point.
    - They hit me
    - My friends say it's ok
    - My parents do it

Please see explanation provided for items 3 and 7.

19. Question 64 – Suggest offering an example (eg, gin) of a 'liquor' in the item stem.

Good point. HBSC permits adding country-specific examples in these questions. We would like to add two examples: '(for example, gin, vodka)'.

20. Question 65 -- Simplify the stem and words in the responses. Perhaps: "Think about the first time you had a drink of alcohol." Also, why does option 5 increase so rapidly from "more than two sips" to "five drinks?" Is that really supposed to be less than "get drunk?"

These are standard questions in the field, used in both the US (Monitoring the Future) and Europe (ESPAD). We are, therefore, reluctant to change the stem because it could threaten the validity of the item and the generalizability of the data. Five drinks has been established as the criterion for 'binge' drinking in adolescents and is considered a particularly high-risk problem behavior. Therefore, it is of greater interest to experts in the field than the continuum of drinking from one to four drinks.

21. Question 66 -- Why provide such a refined division of time? Has this level of precision been previously found reliable?

These are standard questions in the field, used in both the US (Monitoring the Future) and Europe (ESPAD).

22. Question 67--This question is needlessly confusing. Also, how much alcohol consumption provokes drunkenness differs by more than just gender.

This is a standard question used in OMB-approved YRBS. Our pilot subjects had no difficulty with this question.

23. Question 69-- Include "at school?"

This is a standard question asked in ESPAD. We had similar thoughts, but were concerned that schools might object to having this as an option. The 'other places' option permits respondents to write in 'school' without offending school officials.

24. Question 70--Again, choice a is poorly worded and of uncertain use. So is choice e. Other suggestions:

    - "It makes me look cool"
    - "It relaxes me"
    - "My friends give me access to ___"

Please see response to items 3 and 7 above.

25. Add "chew tobacco" as a separate question?

We have included a question about chewing tobacco in previous HBSC surveys but because of the low prevalence and limited interest of investigators it was dropped from HBSC this year and not included in the NEXT survey either.

26. Questions 70 through 73--Where are the response options for those students who do not engage in any of these behaviors? Where is the option for "I don't" or "non applicable?"

See explanation provided for items 3 and 7. During our pilot test of the survey we conducted cognitive interviews with the respondents and they understood that they should indicate the reasons they did or **DID NOT** engage in the behavior.

27. Finding out why people do not engage in these behaviors is likely as powerful and as important as finding out why people do.

We agree. That is why the stem asked them to indicate the reasons they decide whether or not they engage in this behavior.

28. Question 79 --Modernize the question's first two responses to:
    - Receive call
    - Make call

We see your point. These were previously approved by OMB but for clarity we would like to revise these to:

'Receive a call on your cell phone'
'Make a call on your cell phone'

29. Question 80 – What is the demonstrated reliability of these questions? Who will admit to any of the following? Indeed, who will know that they actually perpetrated the action? Who will remember they did so, particularly in the month timeframe? The following appear especially problematic:
    - Changed lanes with very little room between vehicles
    - Pulled out into traffic without waiting for a large space between vehicles – may not realize they did this, so this item captures the adolescents who did this and realized they did this
    - Reading, grooming, or engaging in similar activities – too jumbled, too awkward, too outdated with "grooming"
    - Drinking drugs or using illegal drugs – separate

We and others have used these questions in a number of studies and they are the best available. The validity of self report measures is difficult to assess and the validity of these items and the index have not been established, although we could estimate it by looking at the association of the combined items with outcome measures such as crashes/near crashes and elevated g-force events – however, the data on these outcomes is just becoming available. We have used this measure as a dependent variable in studies evaluating the OMB-approved Checkpoints Program and have found differences after 12 months between groups favoring the Checkpoints Program vs the comparison. Also, in prospective studies we have found that parent limit setting is associated with less self reported risky driving behavior (Simons-Morton et al., 2006a, 2006b).). Risky driving among young drivers is an extremely important health concern and the use of these items in a longitudinal context is will allow for further testing of the validity of these items on a large sample.

30. Questions 82 through 85 -- Repeat these questions for drug use? Also, extensive research at the Census bureau suggests that up to 40% of adults can't accurately answer a similar health insurance question. Suggest dropping or greatly simplifying.

Inclusion of these questions was greatly influenced by two of our funding agencies. The questions on alcohol use were requested by NIAAA and the questions on health insurance are for HRSA. We are trying to keep the survey as short as possible and so included as few questions as necessary. The health insurance question includes a 'don't know' choice. We are interested in what adolescents know and whether this knowledge changes in the first year after high school.

31. Question 89--Include school nurse's office? What is the reason for so many response options likely to be inapplicable to most teens? Suggest simplifying the options for clinic.
    1) Hospital clinic
    2) Doctor's office

Health care utilization is important for HRSA and there are significant differences between seeking care, for example, at an emergency room versus hospital-based clinic. We would like to examine changes in these locations over time.

32. Question 100--Sick vs. retired vs. student… these are all extremely different categories of employment. Separate out for analytical reasons. Also include "He cannot get a job" to capture

those respondents whose fathers may have stopped actively seeking work from despondency or a tough labor market.

We are looking for broad categories for those not working: seeking work, established caretaker role, other (sick, student, retired).

33. Why is there no closure to the survey? How about a "thank you?"

Our health researchers will personally thank the respondent (several times) when they complete the survey. However, we agree and would like to add a large "THANK YOU" to the end of the survey.

Please let me know if you would like us to provide a copy of the survey with the changes we request.

NEXT Plus Parent In-Home Survey

1. Where does it say that the survey is voluntary?

2. Question 1, part a--The exceptions seem unnecessarily burdensome and confusing. Essentially, the item asks for adolescent exposure to cancer, except skin cancer, except melanoma.

This question has been used in previous NIH studies of adolescent health without problems. A health researcher will be present when this survey is being completed and will be able to clarify any confusion.

3. Question 1, part d and f--These two parts are seeking parents to judge the counterfactual. How would parents know what ailment their adolescent would have if pregnant, if the child has never been pregnant?

We would like to modify choices d and f to:
d. High blood pressure or hypertension (when pregnant only – do not answer for boys)
f. High blood sugar or diabetes (when pregnant only– do not answer for boys)

4. How do parents of sons answer these parts to the question?

Please see above.

NEXT Plus Student In-Home Survey

1. Question 1--Is this an exhaustive or inclusive list of aspirin(based) medications? How were these compiled?

These are just examples. In previous work we have been surprised that even elementary school children are very knowledgeable about over-the-counter medications. The reason we are asking these in the home is that adolescents can easily check the medicines that they take if necessary.

2. Question 2--Can the respondents see their medication for reference? How will this be collected?

Yes. The reason we are asking these in the home is that adolescents can easily check the medicines that they take if necessary.

3. There is no transition between questions on medications and on types of facilities in the adolescents' immediate areas. For credibility, there should be a transition.

We would like to bold and increase the font size of the transition that appears on the top of page 4 of the survey.

4. Question 5--The phrase "different facilities" in the item stem will likely confuse adolescents as to what facilities are intended. The entire stem is awkwardly, clumsily worded.

This was not intended as a stem, but rather as the transition between questions on medicine use and on the neighborhood. Question 5, about housing, and its stem appear immediately below this transition statement.

5. Why are townhouses, row houses, apartments, or condos of a certain height all included in the same response option?

Neighborhood assessments, such as this, attempt to differentiate general characteristics of the neighborhood rather than specific differences between row houses, townhouses, or other similar structures. Structures such as those in this general category are quite different from free-standing houses and large apartment complexes but not very different from each other. Furthermore, a row house in one community may be described as a townhouse in another.

6. Question 9--The response to this item/question may vary by the age of the respondent or by the age the respondent is contemplating when reading the question.

This is part of an established measure of neighborhood context. In this question we are more interested in the respondent's perception rather than actual crime statistics. The same neighborhood may even be perceived differently by different members of the same family.

7. Question 11--What can analysts deduce and produce from this question? Drug dealing and the private acts of drunkards may be "interesting things to look at while walking" in a neighborhood but are not necessarily pleasant or healthy to witness.

Again, we are interested in the respondent's perceptions. 'Interesting' is a strong motivator for behavior. Respondents are not likely to refer to crime or threatening behavior as interesting unless it is something that makes their walks interesting. We agree that everything 'interesting' might not be healthful or what we would want these children to see.

8. Question 12--What can analysts deduce and produce from this question? The sidewalks in Georgetown are very cracked and not well-maintained, but that seems like a very safe community relative to downtown where the sidewalks are well-maintained.

This is part of an established scale that assesses overall environment and not only safety. Other questions address safety. It is a combination of factors that contribute to the walkability of a neighborhood.

9. Question 15--Why not have this question about the neighborhood's crime rate during the day follow the question about the crime rate in the neighborhood during the night?

We are not certain why this instrument has been used with the items in this order. We agree that it would make more sense to have these two questions follow each other and would like to move question 15 to follow question 9.

10. Questions 16 & 17--Are these not the same questions as asked earlier? The respondents just answered a question about how the sidewalks are maintained, but these items actually offer an "n/a"

These questions more clearly ask about the presence of these in the neighborhood rather than the condition of these in the neighborhood.

11. Question 18--How are there no roads in a neighborhood? Why are these formatted differently?

Pedestrian-free zones, with residences above street level stores, are very common in Europe. There are increasing numbers of traffic-free zones (pedestrian ways) in urban areas of the US as well and we need to include this option in this longitudinal study.

12. Now the question becomes that of motor vehicles in general? Are scooters, mopeds, and jet-skis counted as motor vehicles?

We agree that question 19 is out of place and would like to move it to follow the current question 20. We provide cars, trucks, etc. as examples but these are not exclusive.

13. Question 19--The four-way intersection question seems like a non-sequitur.

We would like to move question 20 about the 4-way intersection so it follows question 18.

References

Albert PS. Tutorial in biostatistics: Longitudinal data analyses (repeated measures) in clinical trials. Statistics in Medicine. 1999;18:17717-1732.

Baranowski T, Masse LC, Ragan B & Welk G. How many days was that? We're still not sure, but we're asking the question better! Medicine & Science in Sports & Exercise 2008;40:S544-S549.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

Dahlberg, L. L., Toal, S. B., & Behrens, C. B. (1998). *Measuring Violence-Related Attitudes, Beliefs, and Behaviors Among Youths: A Compendium of Assessment Tools*. Atlanta, GA: Center For Disease Control and Prevention, National Center for Injury Prevention and Control.

Drews KL, Harrell JS, Thompson D, Mazzuto Sl, Ford EG, Carter M, Ford DA et al. Recruitment and retention strategies and methods in the HEALTHY study. International Journal of Obesity. 2009;33:S21-S28.

Hu FB, Willett WC, Colditz GA, Ascherio A, Speizer FE, Rosner B, Hennekens CH, Stampfer MJ. Prospective study of snoring and risk of hypertension in women. Am Jo Epidemiology. 1999, October; 150 (8) 806-816.

Little, R. J. A. and Rubin, D. B. *Statistical Analysis with Missing Data*, Wiley, New York, 1987.

Morrison et al., Tracking and follow-up of 16,915 adolescents: minimizing attrition bias. Controlled Clinical Trials, 1997;18:383-396.

Muthén, L. K. & Muthén, B. O. (2001). *Mplus User's Guide* Los Angeles: Muthén and Muthén.

Sassani A, Findley LJ, Kryger M, Goldlust E, George C, Davidson TM. Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome. Sleep. 2004 May 1;27(3):453-8.

Simons-Morton BG, Hartos J, Leaf W, Preusser D. The Effects of the Checkpoints Program on Novice Teen Driving Limits and Outcomes. Accident Analysis and Prevention, 38:907-912, 2006a.

Simons-Morton BG, Hartos J, Leaf W, Preusser D. Do recommended driving limits affect teen-reported tickets and crashes during the first year of teen independent driving? Traffic Injury Prevention, 7:1-10, 2006b.

Sirard JR, & Pate RR. Physical activity assessment in children and adolescents. Sports Medicine, 2001;31:439-454.

Trost SG, Pate RR, Freedson PS, Sallis JF & Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? Medicine and Science in Sports and Exercise 2000;32:426-431.

Villaneuva AT et. al. Sleep Med Rev. 2005 Dec;9(6):419-36. Epub 2005 Sep 22.

Wickel EE, Eisenmann JC, Pangrazi RP, Graser SV, Raustorp A, Tomson LM & Cuddihy TF. Do children take the same number os steps every day? American Journal of Human Biology 2007;19:537-543

Wu, M. C. and Carroll, R. J. (1988) Estimation and comparison of changes in the presence ofinformative right censoring by modeling the censoring process. Biometrics, 44, 175–188.

Young T, Finn L, Peppard PE, Szklo-Coxe M, Austin D, Nieto FJ, Stubbs R, Hla KM. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. Sleep. 2008 Aug 1;31(8):1071-8.