

# The National Longitudinal Study of Adolescent Health Dataset

*-UC DATA Seminar Series: Getting Acquainted  
with Research Data*

*Su Li*

*sli@law.berkeley.edu*

*Statistician of Empirical Legal Studies*

*U.C. Berkeley*

*School of Law*

*April 18<sup>th</sup>, 2012*

# Outline

- Overview of the Add health data
- Participants
- Sampling frame
- Topics the data address and sample size of each component/wave
- Special components of the dataset: Network data; Transcripts data; ONE data
- Tips on analyzing this dataset: Design effects and Weights
- Tips on analyzing this dataset: Non-responses
- How to access the data?

# What is National Longitudinal Study of Adolescent Health?

- **“The National Longitudinal Study of Adolescent Health (Add Health)** is a nationally representative study that explores the causes of health-related behaviors of adolescents in grades 7 through 12 and their outcomes in young adulthood.
- Data collected in 1994-95 (wave I), 1996 (wave II), 2001-02 (wave III), and 2007-08 (wave IV).
- Sample size (varies by wave)
- Participants: sampled adolescents (focus), school administrators, parents, romantic partners etc.

# Participants: adolescents\* +parents+ administrator + others

age in Wave I:13-21 ( in-school questionnaires; in-home interviews; school administrator + parent questionnaires)

age in Wave II: 14-22 (in-home interviews + school administrator phone interviews)

age in Wave III: 18-26 (in-home interviews + DNA sample)

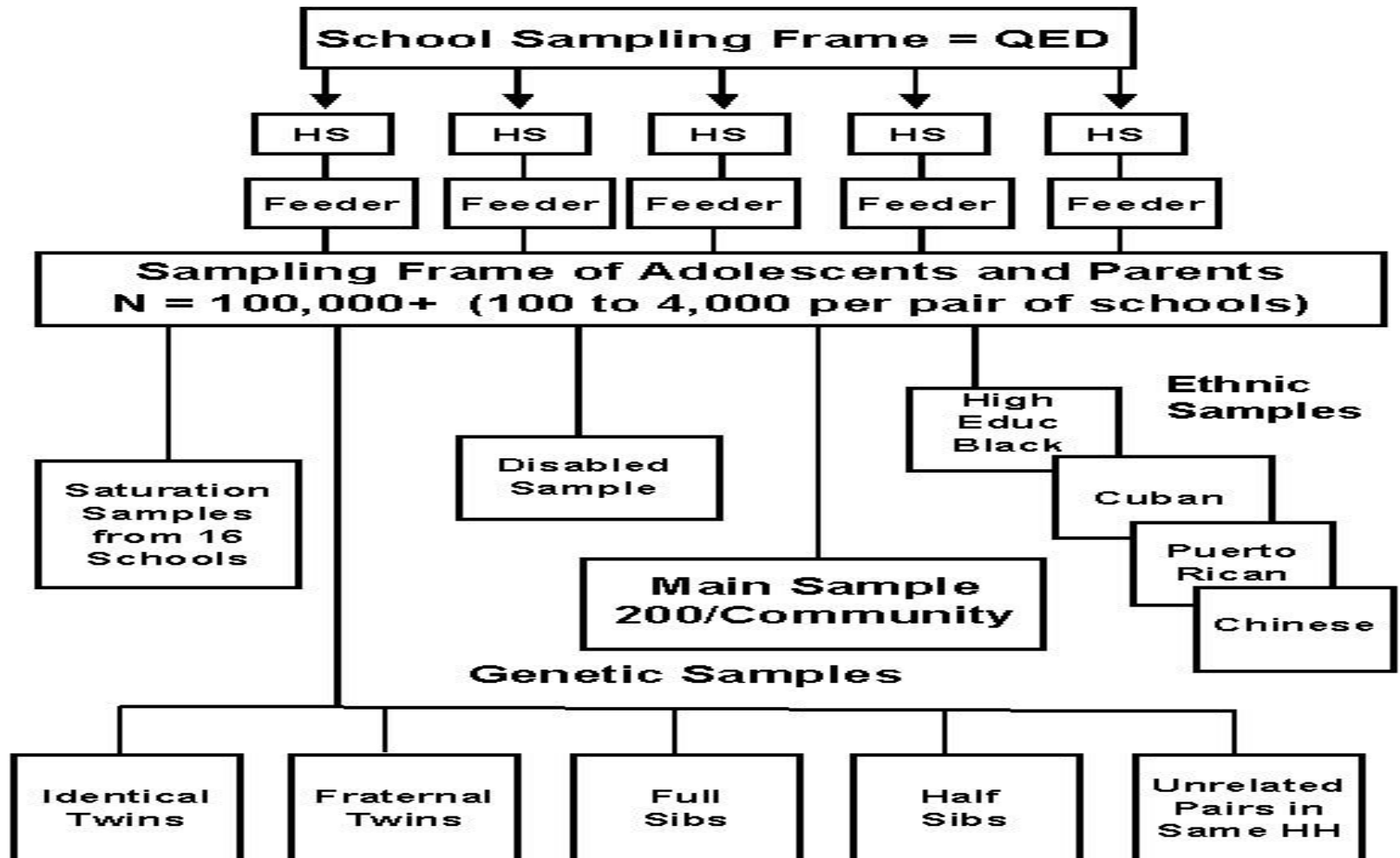
age in Wave IV: 24-32 (in-home interviews + biological sample collection)

Wave I		Wave II					Wave III	Wave III				Wave IV	Wave IV
1994	1995	1996	1997	1998	1999	2000	2001	2002	...	2006	2007	2008	
Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	postHS1	postHS2	3	...	postHS7	postHS8	postHS9	
Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	postHS1	postHS2	postHS3	4	...	postHS8	postHS9	postHS10	
Grade 9	Grade 10	Grade 11	Grade 12	postHS1	postHS2	postHS3	postHS4	5	...	postHS9	postHS10	postHS11	
Grade 10	Grade 11	Grade 12	postHS1	postHS2	postHS3	postHS4	postHS5	6	...	0	postHS11	postHS12	
Grade 11	Grade 12	postHS1	postHS2	postHS3	postHS4	postHS5	postHS6	7	...	1	postHS12	postHS13	
Grade 12	postHS1	postHS2	postHS3	postHS4	postHS5	postHS6	postHS7	8	...	2	postHS13	postHS14	

Sampling Frame: Wave I: 80 high schools were selected from a sample frame of 26,666; Feeder middle schools are located through selected high schools

\*figure from <http://www.cpc.unc.edu/projects/addhealth/data/guides>

**Figure 1. Sampling Structure for Add Health**



# Wave ONE: In-school questionnaires (90,118 adolescents from 144 middle and high schools)

- social and demographic characteristics of respondents
- education and occupation of parents
- household structure
- risk behaviors
- expectations for the future
- self-esteem
- health status
- friendships
- school-year extracurricular activities

# Wave ONE: in-home interviews (20,745 adolescents, oversampled disabled, racial minorities, twins etc) (core sample is 12,105)

- health status
- health-facility utilization
- nutrition
- peer networks
- decision-making processes
- family composition and dynamics
- educational aspirations and expectations
- employment experience
- the ordering of events in the formation of romantic partnerships
- sexual partnerships
- substance use
- criminal activities

# Wave ONE: school administrator (164 administrators) questionnaires and parent questionnaires (17,713- 17,669 parents)

- **School Administrator Questionnaires**

- ❖ school policies and procedures,
- ❖ teacher characteristics,
- ❖ health-service provision or referral
- ❖ student body characteristics.

- **Parent Questionnaire**

- ❖ A parent, preferably the resident mother, of each adolescent respondent interviewed in Wave I was asked to complete an interviewer-assisted, op-scanned questionnaire covering topics such as these:
- ❖ inheritable health conditions
- ❖ marriages and marriage-like relationships
- ❖ neighborhood characteristics
- ❖ involvement in volunteer, civic, and school activities
- ❖ health-affecting behaviors
- ❖ education and employment
- ❖ household income and economic assistance
- ❖ parent-adolescent communication and interaction
- ❖ parent's familiarity with the adolescent's friends and friends' parents



# Wave TWO: in-home interviews (14,738 respondents) + school administrator (128) phone interviews

- **In-Home Interview: Wave II**

- ❖ similar to that at Wave I.
- ❖ **sun exposure and more detailed nutrition questions were added.**
- ❖ Questions about attributes that should not change, such as ethnic background, were not repeated.
- ❖ Physical and functional limitations questions were omitted because the disabled sample was not re-interviewed.

- **School Administrator Telephone Interviews**

- ❖ update information from the first year
- ❖ add information about specific dress codes and security procedures on their campuses.

# Wave THREE Data (Wave I respondents that can be located 15,197)

- In-home interviews (topics are similar to in-home interviews of Wave I and II, with the follow additional information: **Military; Involvement w/criminal justice sys; Children and parenting; Civic participation; Gambling; Mentoring**)
  - ❖ Geocodes for addresses
  - ❖ High school transcripts data
  - ❖ Biological Specimen data (urine; saliva; DNA)
- Opposite sex partner sample (1507 partners) interviews (500 married, 500 cohabiting, 500 dating)
- Binge Drinking sample: A special sub-sample of freshman and sophomores in 2- and 4-year colleges, along with a control group of non-college same-age peers, who were administered additional questions about binge drinking

# Wave FOUR data (15,701 respondents)

- Follow-up of all respondents interviewed in wave I (1994-1995)
  - ❖ Survey/interview data (cover similar topics waves I, II and III, with the following additional information: **Substance use and abuse; Work attitudes and chars; Big 5 Personality, stressors; Cognitive function; Psychosocial factors; Medications** )
  - ❖ Geography data (home address + GPS locations)
  - ❖ Biological data: blood pressure readings, anthropometric measures (height, weight, and waist circumference), saliva for DNA, and blood spots from a fingerstick from all consenting respondents.
- Genome Wide Association Study (GWAS): approximately 12,200 Wave IV saliva. When released, the genetic data and a subset of the phenotypic data will be available through dbGaP (database of Genotypes and Phenotypes)
- Intra-Individual Variation (IIV) Study : 100 Wave IV respondents were selected for repeat collection of biomarkers at a follow-up visit. Purpose of repeated measures is to examine the reliability of biological data.

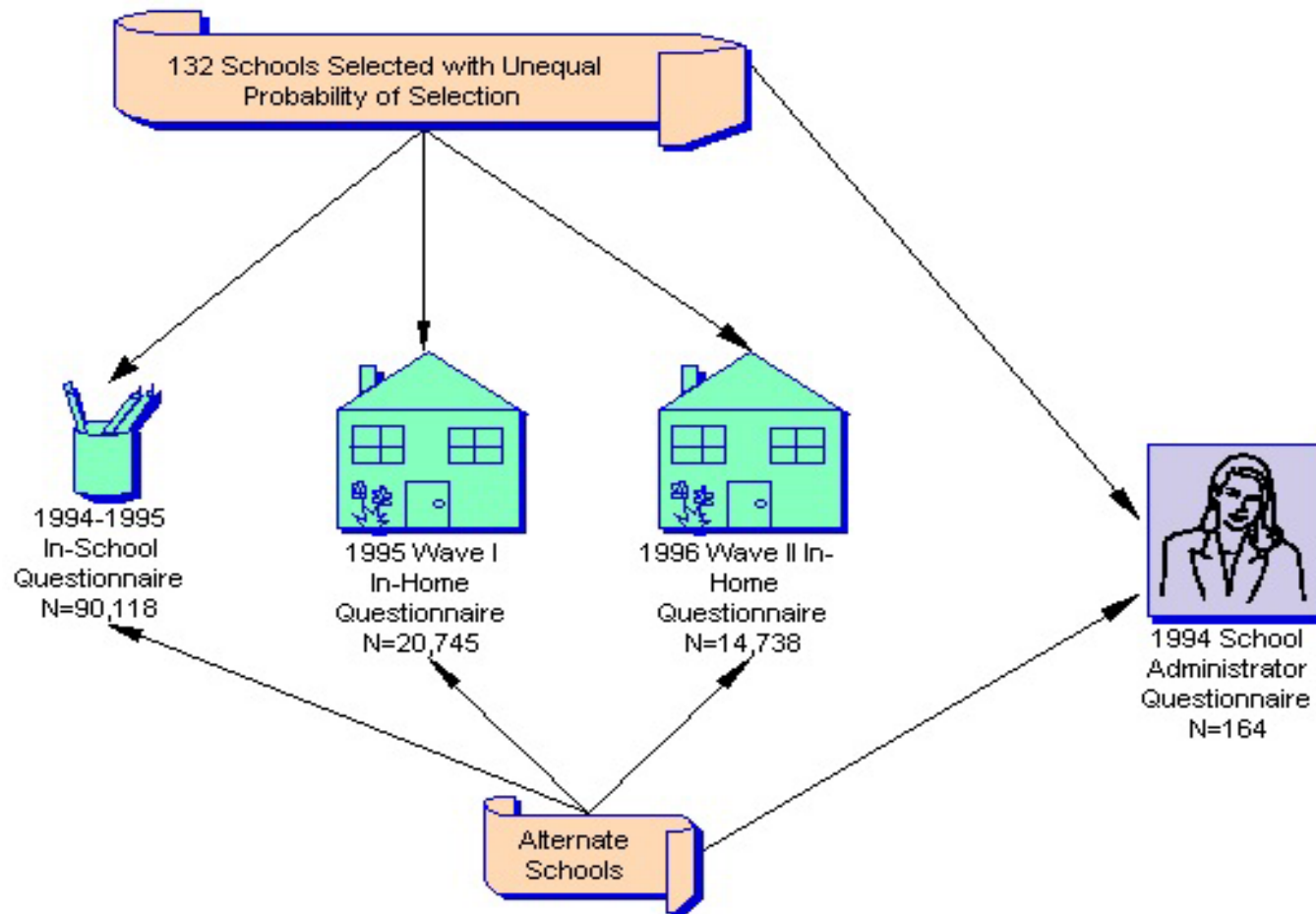
# Special data components

- Friendship Network data (Wave ONE)
  - ❖ 16 saturated samples (high schools), each student were asked to name up to 5 of their best friends. (data in pajek format available)
  - ❖ Constructed network data (network structure variables, such as density, centrality etc.) (Jame Moody)
- High school transcript data(adolescent health and academic achievement study AHAA) (Wave THREE)  
<http://www.laits.utexas.edu/ahaa/>
  - ❖ educational achievement,
  - ❖ course taking patterns
  - ❖ curricular exposure
  - ❖ educational contexts within and between schools
- Obesity and neighborhood environment data (ONE)  
<http://www.cpc.unc.edu/projects/onedata>

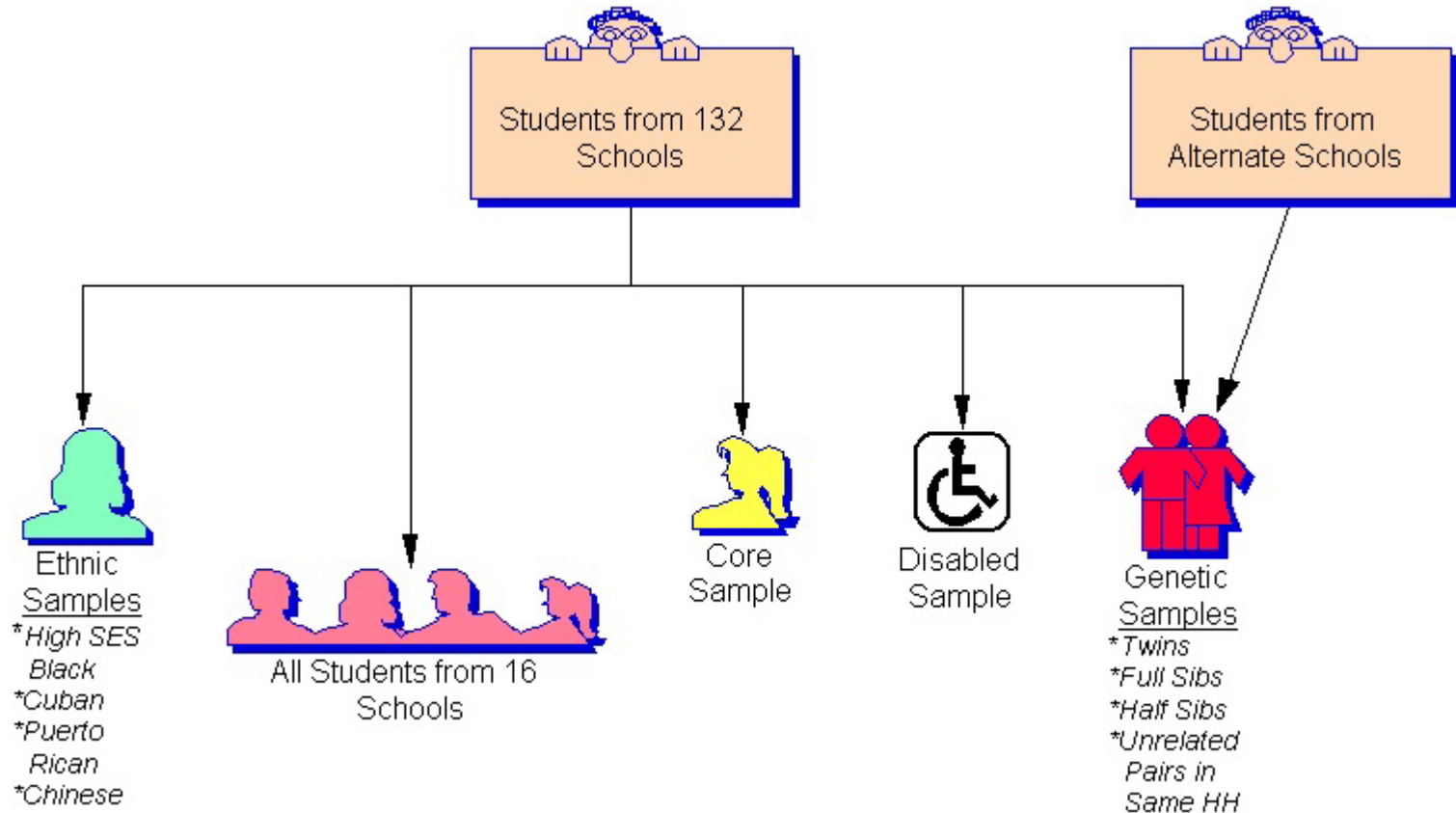
# Tips on data analysis: design effects and weights (Tourangeau & Shin 1999; Chantala & Tabor 1999)

- Special characteristics of add health data (and other complicatedly-designed survey data):
  - design effect (clustering, stratification etc)
    - ❖ sample of 80 high schools (stratified random sample) and 52 middle schools (randomly selected from feeder schools each high school named, only one middle school selected for each high school, some high school do not have feeder schools).
    - ❖ 28 initially selected high school did not cooperate and need to be replaced (on the basis of: school sizes, type, urbanicity, percent white, grade span, census region, etc.)
  - Oversampling: ethnics groups; siblings and twins; disabled samples

# Wave I in-school questionnaire sampling (Chanatala & Tabor 1999)



# Wave I in-home interviews sampling (Chanatala & Tabor 1999)



# Tips on data analysis: design effects and weights (Tourangeau & Shin 1999; Chantala & Tabor 1999)

- Point estimates (means, regression parameters, proportions, etc.) are affected by only the weights
- Variance estimates are affected by the clustering, stratification, weight and design type



# Tips on data analysis: design effects and weights (Tourangeau & Shin 1999; Chantala & Tabor 1999)

- Recommendation: always use the survey analysis software, such as Stata
- With replacement as the design type
- Svy commands of stata; use region as stratum, schools as PSU or cluster var.
- And apply grand sample weight (have combined weights for all waves of data) (eliminate cases with missing values of weight from your analysis)
- Note: even if you are only interested in a sub-population, you need to start from the whole data to reduce errors in standard error.
- If use multiple panel data in one analysis, use the weight of the most recent wave data.
- See Chantala & Tabor 1999 for a step-wise guide

# Tips on data analysis: non-respondents

- Wave Three data: “Wave III sample adequately represents the same population as the Wave I sample when final sampling weights are used to compute population estimates.” (Chantala et al 2003 )
- Wave Four data: “Wave IV non-response bias is negligible and the Wave IV sample adequately represents the same population surveyed at Wave I.” (Chantala et al 2008)
  - Wave Four Exception, e.g. 35% relative bias for the lowest group with APHVT scores less than 70.

It may mean significantly more non-responders than responders did in fact have very low AHPVT scores at Wave I, but may also be caused by the small number of respondents that are in this category.

# How to access the data

- Public-use data: download from ICPSR or obtain CD from sociometrics

<http://www.cpc.unc.edu/projects/addhealth/data/publicdata>

- Restricted-use data

-obtain your own data:

1. IRB approval
2. Data application through ICPSR + \$800 fee

-join an existing data contract upon approval of PI

1. IRB approval
2. Participant application through ICPSR
3. Follow the security protocols and share the data access upon ICPSR approval

# other

- Add health conference, July, biennially

<http://www.cpc.unc.edu/projects/addhealth/news/call-for-papers-2012-add-health-users-conference>

# Focuses of topics at each wave

- Waves I and II: forces that may influence adolescents' health and risk behaviors, eg. personal traits, families, friendships, romantic relationships, peer groups, schools, neighborhoods, and communities.
- Wave III, how adolescent experiences and behaviors are related to decisions, behavior, and health outcomes in the transition to adulthood
- Wave IV: developmental and health trajectories across the life course of adolescence into adulthood using an integrative approach that combines the social, behavioral, and biomedical sciences in its research objectives, design, data collection, and analysis.

# References

- Harris, K.M., C.T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J.R. Udry. 2009. The National Longitudinal Study of Adolescent Health: Research Design/ [WWW document]. URL: <http://www.cpc.unc.edu/projects/addhealth/design>
- Tourangeau, R & Shin, H. 1999 Grand Sample Weight  
<http://www.cpc.unc.edu/projects/addhealth/data/guides/weights.pdf>
- Chantala, K. & Tabor J. 1999 *Strategies to Perform a Design-Based Analysis Using the Add Health Data*  
<http://www.cpc.unc.edu/projects/addhealth/data/guides/weight1.pdf>
- Chantala, K. (2003). [Analyzing Add Health Data](#). Add Health Users Workshop. Bethesda, MD.
- Chantala, K. (2003). [Using Weights with Wave III Data](#). Add Health Users Workshop. Bethesda, MD.
- Chantala, K. (2004). [Guidelines to Analyzing Add Health Data: Part 1](#). Add Health Users Workshop. Bethesda, MD.
- Chantala, K. (2004). [Guidelines to Analyzing Add Health Data: Part 2](#). Add Health Users Conference. Bethesda, MD.

- -end

# Topics

- Physical and mental health, such as weight and height, injury and disability, dietary patterns and physical activity, substance use, access to and use of health care services, and suicide and depression
- Interpersonal relationships and sexual behaviors, such as family relationships, friendships, interracial relationships, faith community interactions, sexual activity, and sexual orientation
- Education, including cognitive ability and individual, family, peer, and community influences on school performance
- Delinquency and violence, including individual, family, peer, and community influences on delinquency and violence and risk factors for delinquency and violence
- Involvement in adult roles, including parenthood, jobs, marriage
- Genetic characteristics and biological measures that indicate the presence of specific diseases and disease processes
- Measures of the environments in which participants live and go to school