Add Health Research Design
Waves I-IV
Sampling Structure

School Sampling Frame = QED

Saturation Samples from 16 Schools

Disabled Sample

Main Sample 200/Community

Genetic Samples

Identical Twins

Fraternal Twins

Full Sibs

Half Sibs

Unrelated Pairs in Same HH

Sampling Frame of Adolescents and Parents N = 100,000+ (100 to 4,000 per pair of schools)

Ethnic Samples

High Educ Black

Puerto Rican

Chinese

Cuban

Sampling Frame of Adolescents and Parents N = 100,000+ (100 to 4,000 per pair of schools)
Longitudinal Design

Wave I 1994-1995
- Students: 90,118
- School Admin: 144

Wave II 1996
- School Admin: 128

Wave III 2001-2002
- Adolescents in grades 7-12: 20,745
- Parents: 17,670

Wave IV 2007-08
- Adolescents in grades 8-12: 14,738
- Young Adults Aged 18-26: 15,197
- Partners: 1,507
- Adults Aged 24-32
Wave I Sample Design

- 80 high schools were selected from a sample frame of 26,666.
- Prior to sampling, schools were sorted by:
  - size
  - school type
  - census region
  - level of urbanization
  - percent white
Replacement Schools

- Of the 80 selected high schools, 52 were eligible and agreed to participate.
- The remaining 28 schools were replaced by similar high schools. They were found by sorting the frame by eight variables:
  - school size
  - school type
  - level of urbanization
  - percent white
  - grade span
  - percent black
  - census region
  - census division
Replacement Schools

- Schools were sorted in a random order within each category.
- The replacement school was the school that followed the initial sample school on the sorted file.
- As a result, the replacement school matched the selection it was replacing with respect to the eight characteristics.
Replacement Schools

• If the first replacement school was ineligible or failed to cooperate, it was replaced by the next school on the list.

• Within some categories, there were not enough potential replacements, or the list of replacements was exhausted. In these cases, similar categories were combined and the file resorted.
Feeder Schools

- Participating high schools were asked to identify junior high or middle schools that were expected to provide at least 5 students to the entering class of the high school.

- A single feeder school was selected for each high school.
Feeder Schools

• The feeder’s probability of selection was proportional to the percentage of the high school’s entering class that came from that feeder.

• Four schools had no eligible feeder, as students came from a very large number of junior high or middle schools.

• Twenty high schools were their own feeder schools, as they had grade ranges that included 7th or 8th grades.
### Participating High Schools

<table>
<thead>
<tr>
<th>Region</th>
<th>Grade Span</th>
<th>Metropolitan Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>K-12</td>
<td>Urban</td>
</tr>
<tr>
<td>South</td>
<td>7-12</td>
<td>Suburban</td>
</tr>
<tr>
<td>Midwest</td>
<td>9-12</td>
<td>Rural</td>
</tr>
<tr>
<td>West</td>
<td>10-12</td>
<td></td>
</tr>
</tbody>
</table>

### School Type

<table>
<thead>
<tr>
<th>School Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>71</td>
</tr>
<tr>
<td>Catholic</td>
<td>3</td>
</tr>
<tr>
<td>Other Private</td>
<td>6</td>
</tr>
</tbody>
</table>
Parental Consent

• Parental consent was required to list student names in a directory and to allow students to participate in the study.

• Unless otherwise directed by the school, passive consent forms were used (it was assumed that a parent granted permission unless the form was returned with a signature that indicated otherwise).

• Some schools required active consent forms (the form had to be returned with a signature indicating that permission was granted).
Security System

- To protect the identities of participants, a rigorous security system prevents anyone from being able to link a respondent’s answers to a name or other identity.
- Identification numbers used to collect data are never used for data distribution.
- At the same time, the security system allows researchers to link questionnaires across all components of the study.
In-school Components

- 145 middle, junior high, and high schools participated.
- 90,118 students completed a 45-minute questionnaire.
- Each participating school was asked to complete a School Administrator questionnaire.
In-school Questionnaire

Gathered general descriptive information about the student:

• student’s and parent’s background
• his or her friends
• school life
• school work and school activities
• general health status and health-related behaviors
School Administrator Questionnaire

Gathered information about the educational setting and environment of the school.

- Content areas included:
  - general characteristics of the school and student body
  - curriculum, school services, and programs
In-home Components

- Each school provided a roster of all students enrolled.
- From the rosters and the pool of participants in the in-school survey, adolescents in grades 7 to 12 were sampled to participate in the in-home interview.
- Adolescents were interviewed at two points in time, first at Wave I and then a year later, at Wave II.
In-home Interview

• Written informed consent was obtained from the parent or legal guardian and the adolescent.

• A Computer-Assisted Personal Interview (CAPI)/Audio Computer-Assisted Self Interview (ACASI) was administered to the adolescent.

• Sections with more sensitive questions were asked in the self-administered portion of the interview.
Parent Interview

- A parent or guardian was interviewed during Wave I of the study.
- This interview provided further information about the family composition and the adolescent’s health history.
- The questionnaire asked demographic and health-related information about the parent or guardian and general questions about the adolescent respondent.
In-home Sample

• a core sample of adolescents derived from rosters supplied by the selected schools and from in-school participation

• a sample for a study of relationship patterns where all respondents were selected from two schools

• a genetic sample composed of siblings and twins

• a sample of unrelated adolescents who reside in the same household
In-home Sample

- an oversample of black adolescents with college-educated parents
- an oversample of Cuban and Puerto Rican adolescents
- an oversample of Chinese adolescents
- an oversample of physically disabled adolescents (Wave I only)
Sample Selection: Core and PAIRS Samples

• The core sample consisted of roughly equal-sized samples drawn from 12 student-level strata.

• Strata were formed by cross-classifying students by their sex and grade.

• Overall sample-size targets were set for each stratum by dividing the total size of the core sample by the number of strata.
Sample Selection: Core and PAIRS Samples

- School-level targets were also set for each stratum by dividing the overall stratum target by the number of schools with at least one student in the stratum.
- The main frame for selecting the core sample was the set of rosters developed at the sample high schools and their linked feeders.
- At the two (purposively selected) PAIRS schools, all of the students were selected for data collection.
Sample Selection: Non-Genetic Supplements

Eligibility was determined by race/ethnicity and by disability status:

- **High education blacks** – included black students, either of whose parents was a college graduate
- **Cubans** – included students who were from Cuba or whose parents were from Cuba
Sample Selection: Non-Genetic Supplements

- **Puerto Ricans** – included students of Puerto Rican descent
- **Chinese** – included students of Chinese descent
- **Disabled** – included students who had difficulty using their limbs for the year prior to the survey, and, as a result, used a cane, wheelchair, orthopedic shoes, artificial limb, or some other mechanical aid (Wave I only)
Sample Selection: Genetic Supplements

- **Twins** – any student who identified himself as a twin (in the in-school questionnaire); previously unreported twins discovered during the in-home interview were added to the sample at that time.
- **Other siblings of twins** – the non-twin siblings of those in the twin sample; to be eligible they had to be enrolled in grades 7 to 12 at the time of sample selection.
Sample Selection: Genetic Supplements

• **Other full siblings** – full siblings where neither member was a twin and both were enrolled in grades 7 to 12.

• **Half-siblings** – pairs of half-siblings of which both members were enrolled in grades 7 to 12.

• **Non-related** – adolescents living in the same household who did not share the same biological mother or father and were enrolled in grades 7 to 12.
### Wave I In-home Adolescent Participants

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>20,745</td>
</tr>
<tr>
<td>Core</td>
<td>12,105</td>
</tr>
<tr>
<td>Disabled</td>
<td>957</td>
</tr>
<tr>
<td>High ed blacks</td>
<td>1,547</td>
</tr>
<tr>
<td>Cuban</td>
<td>538</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>633</td>
</tr>
<tr>
<td>Chinese</td>
<td>406</td>
</tr>
<tr>
<td>Twins</td>
<td>1,534</td>
</tr>
<tr>
<td>Full-siblings</td>
<td>2,500</td>
</tr>
<tr>
<td>Half-siblings</td>
<td>848</td>
</tr>
<tr>
<td>Non-related</td>
<td>1,314</td>
</tr>
<tr>
<td>Pairs</td>
<td>2,553</td>
</tr>
</tbody>
</table>
Wave II Sample Selection

- The Wave II sample was drawn primarily from the pool of participants in Wave I.
- The majority of 12th-grade respondents were removed from the Wave II sample, as they exceeded the grade eligibility requirement.
Wave II Sample Selection

- Twelfth-graders who were part of a genetic pair were retained.
- Wave I disabled sample was not re-interviewed at Wave II.
- The Wave II sample contains a small number of adolescents who did not participate in the first wave.
- No parent interview was conducted during the second wave.
Wave III In-home Components

- **Wave III respondents** 15,197
  - Includes 15,170 original Wave I respondents and 27 Wave II special genetic respondents
  - 14,979 were interviewed during the main study and 218 were interviewed during the pretest

- **Romantic partners** 1,507
  - Partners of Add Health respondents, new in Wave III
Wave III Interview

- Interviewer administered in-home interview (CAPI) with sections containing sensitive questions asked in the self-administered portion.
- Respondents who agreed to provide a urine and/or saliva sample were asked to read and sign an additional consent form.
- Respondents received an additional incentive for these samples.
Wave III Biomarkers

- Urine was analyzed for the presence of three sexually-transmitted diseases (STDs), chlamydia, gonorrhea, and trichomoniasis.
- Saliva was collected and analyzed for DNA from all full sibs and twins interviewed at Wave III.
- Additional saliva was analyzed for the presence of HIV antibodies.
Wave III Binge Drinking

• All respondents with Wave I and II data who were in the 7th or 8th grade at Wave I were pre-identified to answer questions about attitudes toward binge drinking.

Composition

• 700 attending two or four year college: 350 males, 350 females, never married.

• 700 not attending college: 350 males, 350 females, never married.
Wave III Couples

- 1500 from pre-selected group of 10,000
- final quota to include 1/3 married, 1/3 cohabiting, 1/3 dating
- heterosexual couples only
- 18 years or older
- current relationship
- relationship had lasted three months or more
Additional Wave III Components

• measured and self-reported height and weight
• latitude and longitude of current residence
• questions about contact with school friends
• information on relationships with siblings
Wave III Sample Selection

- The Wave III sample was drawn primarily from the pool of participants in Wave I.
- Respondents were required to be 18 years of age or older.
- Partners had to be 18 years old or older, opposite sex, in a current relationship that had lasted for at least three months.
Wave III Consent

- Respondent was asked to read and sign an informed consent form.
- Parental consent was not needed as respondents were 18 to 26 years old.
- All respondents who agreed to participate in the interview received an incentive payment.
Wave IV Sample Selection

- Follow all Wave I respondents.
- Data collection 2007-08 when cohort members are 24-32 and completing transition to adulthood.
- 85% response rate is budgeted.
- N=17,000
Wave IV Builds on Add Health Strengths

- Bring together rich social and behavioral data with biological data relevant to the current and future health concerns of Add Health cohort at ages 24-32.
- Expand collection of biological data to understand genetics, stress, and predisease pathways.
- Focus on obesity, stress, and health risk behavior.
- Trans-disciplinary program of research in Wave IV: “Social, Behavioral, and Biological Linkages.”
Integrative Life Course Theoretical Framework

CONTEXT

BEHAVIOR

BIOLOGY

HEALTH AND WELL-BEING

Childhood → Adolescence → Emerging Adulthood → Young Adulthood
Data to be Collected at Wave IV

- Social/economic status, behaviors, and experiences
- Psychological status and experiences
- Reported health status and health behavior
- Environmental contexts
- Biomarkers including blood pressure, waist circumference, lipids profile, and other metabolic, inflammatory, and immune function measures as well as repeated measures of height and weight
- DNA from saliva samples
Domains of Wave IV Biological Measures

- **Anthropometric:** height, weight, BMI, waist circumference
- **Cardiovascular:** blood pressure, pulse
- **Metabolic processes:** lipids, glucose, glycosylated hemoglobin
- **Immune function:** EBV
- **Inflammatory processes:** CRP
- **Genetic:** 10 candidate loci
Wave IV Biomarker Collection Sequence

- Anthropometric (height, weight, waist)
- Blood pressure
- Finger prick for blood spots on filter paper
- Saliva collected for DNA
Wave IV Intra-Individual Variation (IIV) Study

- Estimate reliability of biomarker measures by collecting repeat measures on 100 respondents

- Interview IIV respondents twice, 1-2 weeks apart
  - Visit 1: full interview + biomarkers
  - Visit 2: abbreviated interview + biomarkers

- Process the biomarkers (labs and technicians masked)

- Use reliability estimates to:
  - Help monitor/assure/control biomarker data quality
  - Correct for measurement error