The National Longitudinal Study of Adolescent Health (Add Health) Twin Data

Kathleen Mullan Harris,[1,2] Carolyn Tucker Halpern,[1,3] Andrew Smolen,[4] and Brett C. Haberstick[4]

[1] Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina, United States of America

[2] Department of Sociology, University of North Carolina at Chapel Hill, North Carolina, United States of America

[3] Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill, North Carolina, United States of America

[4] Institute for Behavioral Genetics, University of Colorado at Boulder, Colorado, United States of America

This article describes the design and data availability for samples of genetic pairs in the National Longitudinal Study of Adolescent Health (Add Health). Add Health provides unique samples of genetic pairs that are nationally representative and followed longitudinally from early adolescence into young adulthood with 3 in-home interviews and a 4th interview planned for 2007 to 2008. The design of Add Health included an embedded genetic sample of more than 3000 pairs of individuals with varying genetic resemblance, including monozygotic twins, dizygotic twins, full siblings, half siblings, and unrelated siblings who were raised in the same household. Add Health has collected rich longitudinal social, behavioral, and environmental survey data, as well as buccal cell DNA from a subsample of the genetic sample (N = 2612). Add Health has an enlightened dissemination policy and to date has released phenotype and genotype data to more than 3000 researchers in the scientific community.

The National Longitudinal Study of Adolescent Health (Add Health) is an ongoing study of a nationally representative sample of more than 20,000 individuals that began with in-school questionnaires administered to adolescents in grades 7 to 12 in the United States in 1994 and 1995. The in-school survey was followed by three waves of in-home interviews in 1995, 1996 and 2001, 2002, and a fourth wave is planned for 2007, 2008, when the sample will be aged between 24 and 32. Add Health provides unique samples of twins that are nationally representative and followed longitudinally from early adolescence into young adulthood. The design of Add Health included an embedded genetic sample of more than 3000 pairs of individuals with varying genetic resemblance, including monozygotic (MZ) twins, dizygotic (DZ) twins, full siblings, half siblings, and adolescents with no biological relationship but who were raised in the same household.

Add Health was developed in response to a mandate from the United States Congress to fund a study of adolescent health and was designed by a nationwide team of multidisciplinary investigators from the social, behavioral, and health sciences. The original purpose of Add Health was to understand the causes of adolescent health and health behavior with special emphasis on the forces that reside in the multiple contexts of adolescent life. Innovative features of the research design facilitated this purpose by providing independent measurements of the social environments of adolescents, including contextual data on the family, neighborhood, community, school, friendships, peer groups, and romantic relationships. Data were gathered from adolescents themselves, their parents, siblings, friends, romantic partners, fellow students, and school administrators. Existing databases with information about the neighborhoods and communities of the adolescents were merged with the Add Health data.

The Add Health cohort was then followed during their transition to adulthood; research questions turned to the meaning of adolescence for trajectories into young adulthood and how adolescent experiences and behaviors are related to social, behavioral, and health outcomes in the transition to adulthood. Across all interview waves, comprehensive data on health and health-related behavior were collected, including diet, physical activity, health service use, morbidity, injury, violence, sexual behavior, contraception, sexually transmitted infections, pregnancy and childbearing, suicidal intentions and thoughts, substance use and abuse, and delinquency. Data were also collected on such attributes as height, weight, pubertal development, mental health status, and chronic and disabling conditions. Achievement and academic progress were updated in each wave, including school performance, school problems, relationships with students and

teachers, self-esteem, self-confidence, and future expectations for education, health, and family formation.

The study has been funded by three Program Projects over the period 1994 to 2010 by the National Institutes of Child Health and Human Development (P01 HD031921) with cofunding from 17 other federal agencies. Add Health has become a national data resource for over 3000 Add Health researchers who have obtained more than 200 independently funded research grants and have produced hundreds of research articles published in multiple disciplinary journals and research outlets. Below we describe the design of Add Health and its twin samples, followed by the type of data available, a summary of research findings, and data access.

Add Health Design and Data

Add Health used a school-based design. The primary sampling frame was derived from the Quality Education Database (QED). From this frame a stratified sample of 80 high schools was selected (defined as schools with an 11th grade and more than 30 students) with probability proportional to size. Schools were stratified by region, urbanicity, school type (public, private, parochial), ethnic mix, and size. For each high school selected, a feeder school was identified and recruited (typically a middle school) with probability proportional to its student contribution to the high school, yielding one school pair in each of 80 different communities. More than 70% of the originally selected schools agreed to participate in the study. Replacement schools were selected within each stratum until an eligible school or school-pair was found. Overall, 79% of the schools contacted agreed to participate in the study. As some schools spanned grades 7 to 12, the sample contains 132 schools, each associated with one of 80 communities. School size varied from fewer than 100 students to more than 3000 students. Add Health communities were located in urban, suburban, and rural areas of the country.

From September 1994 to April 1995, in-school questionnaires were administered to students in these schools. Each school administration occurred on a single day within one 45- to 60-minute class period. Add Health collected in-school questionnaires from over 90,000 students. The in-school questionnaire provided measurement on the school context, friendship networks, school activities, future expectations, and a variety of health conditions. An additional purpose of the school questionnaire was to identify and select special supplementary samples of individuals in rare but theoretically crucial categories. It is this aspect of the design that enabled Add Health to oversample twins (and other genetically-related siblings) within this nationally representative sampling frame.

Add Health obtained rosters of all enrolled students in each school. From the union of students on school rosters and students not on rosters who completed in-school questionnaires, a sample of adolescents was chosen for a 90-minute in-home interview constituting the Wave I in-home sample. To form a core sample, students were stratified in each school by grade and sex and approximately 200 adolescents were sampled from each pair of schools. The core in-home sample is essentially self-weighting, and provides a nationally representative sample of 12,105 American adolescents in grades 7 to 12. Twins and other sibling pairs occur naturally in the core in-home sample proportional to their representation in the general population of adolescents in grades 7 to 12 in the United States in 1995. However, to increase the potential number of twins for genetic analysis, twins were oversampled for the in-home sample based on responses from the in-school survey.

Twin Recruitment

From answers provided on the in-school survey, Add Health drew supplemental samples based on the genetic relatedness of siblings in a household and according to adoption status. If an adolescent indicated that he or she was a twin in the in-school survey, they were selected with certainty (e.g., 100%) for inclusion in the in-home sample. Full siblings occur naturally in large numbers in the core in-home sample, but half-siblings, adopted adolescents, and unrelated adolescents (e.g., stepsiblings, foster children) who live in the same household were also oversampled. Table 1 shows the number of pairs of adolescents in the embedded genetic sample at Wave I. These genetic pairs data include more than 3000 pairs of adolescents who have varying degrees of genetic relatedness and represent a fully articulated behavioral genetic design. These pairs of adolescents took the same questionnaires, share the same home environment, and share, in most cases, the same school and neighborhood environment. The embedded genetic design in Add Health and the data available for genetic pairs are unprecedented for a national study of this magnitude. In all follow-up interviews, high priority has been placed on locating and reinterviewing pairs in the genetic sample to maintain the integrity of this sample for longitudinal research purposes.

In addition to the genetic oversamples, supplemental samples were also drawn based on ethnicity (Cuban, Puerto Rican, and Chinese) and disability. Add Health also oversampled African American adolescents with highly educated parents. Finally, a special 'saturated' sample was included in Wave I by selecting all enrolled students from two large schools and 14 small schools for in-home interviews. Complete social network data were collected in the saturated field-settings by generating a large number of romantic and friendship pairs for which both members of the pair had in-home interviews. These data provide unbiased and complete coverage of the social networks and romantic partnerships in which adolescents are embedded. The core sample plus the special samples produced a sample size of 20,745 adolescents in Wave I. The Wave I in-home sample is the

Kathleen Mullan Harris, Carolyn Tucker Halpern, Andrew Smolen, and Brett C. Haberstick

Table 1

Pairs of Individuals in Embedded Genetic Sample, Add Health Wave I

Sibling type	N
Twin	784
Monozygotic	307
Dizygotic	452
Undetermined twin type	25
Full sibling	1251
Half sibling	442
Nonrelated sibling	662
Adopted*	560

Note: *Single individuals.

basis for all subsequent longitudinal follow-up interviews, and thus this innovative design remains a major strength of the longitudinal data as well.

Longitudinal Design and Data

Table 2 presents the various survey components administered across the three waves of data collection in Add Health, dates of data collection, the number of completed interviews for the entire sample, and the response rates for the prospective longitudinal cohort. The table also shows the projected number and response rate for the planned Wave IV in 2007 to 2008.

A parent, usually the resident mother, also completed a 30-minute interviewer-assisted interview at Wave I. The parent questionnaire gathered data on such topics as heritable health conditions, marriage and marriage-like relationships, involvement in volunteer, civic, or school activities, health-related behaviors, education, employment, household income and economic assistance, parent-adolescent communication and interaction, the parent's familiarity with the adolescent's friends and friends' parents, and neighborhood characteristics.

In 1996, all adolescents in grades 7 through 11 in Wave I *and* 12th graders who were part of the genetic sample and the adopted sample were followed up 1 year later for the Wave II in-home interview ($N = 14,738$), thereby maintaining the integrity and size of the genetic sample. The in-school and Waves I and II in-home interviews constitute the adolescent period in Add Health and contain unique data about family context, school context, peer networks, spatial networks and community context. School context data come from the in-school surveys based on the census of students in each school, as well as from school administrator questionnaires. Peer network data were obtained in the in-school questionnaire. Adolescents nominated their five best male and five best female friends from the school roster (using a unique ID). Because nominated school friends also took the in-school interview, characteristics of respondents' peer networks can be constructed by linking friends' data from the in-school questionnaire and constructing variables based on friends' actual responses. In the in-home Wave I and Wave II interviews, respondents nominated their best friend, as well as their romantic and sexual partners. If their friend or partner was also a member of the in-home sample, their data could be linked to construct friendship and partner contexts. In the 16 schools that were part of the 'saturated' sample, all students in the school were also interviewed in the home. Complete friendship and sexual networks could therefore be constructed with these data.

Spatial data indicating the exact location of all households in the survey were collected using hand-held global positioning system (GPS) devices or recording actual addresses. These data make possible the interweaving of spatial and social networks, and the construction of community contexts. More than 2500 attributes for community and neighborhood contexts at multiple spatial units of observation have been obtained and merged with the Wave I and Wave II survey data to describe the neighborhood and community contexts in which adolescents are embedded. Neighborhood and community data were gathered from a variety of sources, such as the US Census, the Centers for Disease Control and Prevention, the National Center for Health Statistics, the Federal Bureau of Investigation, and the National Council of Churches.

Table 2

Waves of Add Health Data Collection and Interviews

Interview wave (date)	Interview	N	Response rate
Wave I (1994–1995)	In-school adolescent	90,118	
	School administrator	144	
	In-home adolescent	20,745	80%
	In-home parent	17,670	
Wave II (1996)	In-home adolescent	14,738	90%
	School administrator	128	
Wave III (2001–2002)	In-home young adult	15,197	77%
	Partner	1507	
Wave IV (2007–2008)	In-home young adult	17,000*	85%*

Note: * Projected.

At Wave III, all original Wave I in-home adolescent respondents were recontacted and reinterviewed in 2001 to 2002 with a response rate of 77%. This interview captured the Add Health cohort during their transition to adulthood when they were aged 18 to 26 years. In Wave III, quota samples of 500 partners each in married, cohabiting, and dating couples were recruited by Add Health respondents, resulting in the partner sample of 1507. This sample contains partners who both completed the same Wave III interview. Spatial data were again attached to the Wave III individual-level data using the geocodes of the home residence. New data on family formation, college attendance and college context, mentoring, and civic participation were collected at Wave III, while maintaining the longitudinal integrity of previous data on family, friends, achievement, physical and mental health, and health behavior. Also new at Wave III was the collection of several biospecimens. Urine and saliva were collected to test for sexually transmitted diseases (STD) and human immunodeficiency virus (HIV), and buccal cell DNA was collected from the twins and full siblings in the genetic sample (genetic data described in more detail below).

The design structure and data content of Add Health are detailed here for the entire sample because these aspects also describe the Add Health twin samples and overall genetic sample. That is, the twins and all genetic pairs in Add Health have data from a parent, were followed up at Wave II in adolescence, at Wave III during the transition to adulthood, and will be followed-up at Wave IV. Twins and all genetic pairs have school context, peer network, family, partner and relationship, and spatial data.

Table 3 shows the number of completed in-home interviews at Wave I and response rates at Waves II and III for pairs of siblings in the genetic sample. Response rates are based on pairs who were interviewed at Wave I and who were eligible for inclusion in the Wave II sample, and subsequently in the Wave III sample. Response rates for the genetic sample are quite high, especially at Wave II, only 1 year following Wave I, when 95% of the Wave I pairs were reinterviewed.

Even at Wave III, 6 to 7 years after Wave I, response rates are over 85% for all types of genetic pairs, 89% for the entire genetic sample, and 92% and 89% for MZ and DZ twin samples, respectively. Sampling weights for use with the genetic sample are being constructed for individuals and pairs in the genetic sample. Poststratification adjustments will ensure that these weights are calibrated to match distributions available from the current population survey on education, marital status, and race of the household. These weights will be available for use in early 2007.

Figure 1 shows the developmental stages and time period during which the Add Health cohort was followed from early adolescence into young adulthood, and the types of environmental data available across waves of interviews. The strength of Add Health, and thus the strength of the Add Health twin data, is in the multiple levels of environmental data that allow researchers to examine both environmental and genetic effects and to explore gene × environment interactions across the life course from early adolescence into young adulthood.

Phenotype Data

Add Health is an omnibus study that attempted to measure all the domains relevant to the specific developmental stage of the Add Health cohort at the time of the interview, with a particular focus on causes and consequences of health and health behavior. Table 4 lists the questionnaire sections of the Wave I in-home interview that represent the topical areas covered in this first adolescent instrument. Table 5 presents the questionnaire sections of the Wave III in-home interview (Wave II questionnaire sections are similar to Wave I). The Wave I and II interviews collected data relevant to adolescence, including relationships with parents, siblings and friends, academics and school, romantic and sexual relationships, mental health, expectations for the future, and health risk behavior. At Wave III during the transition to adulthood, attention shifted to capture early family formation, postsecondary education, labor market activity and military service, mentoring, civic participation, as well
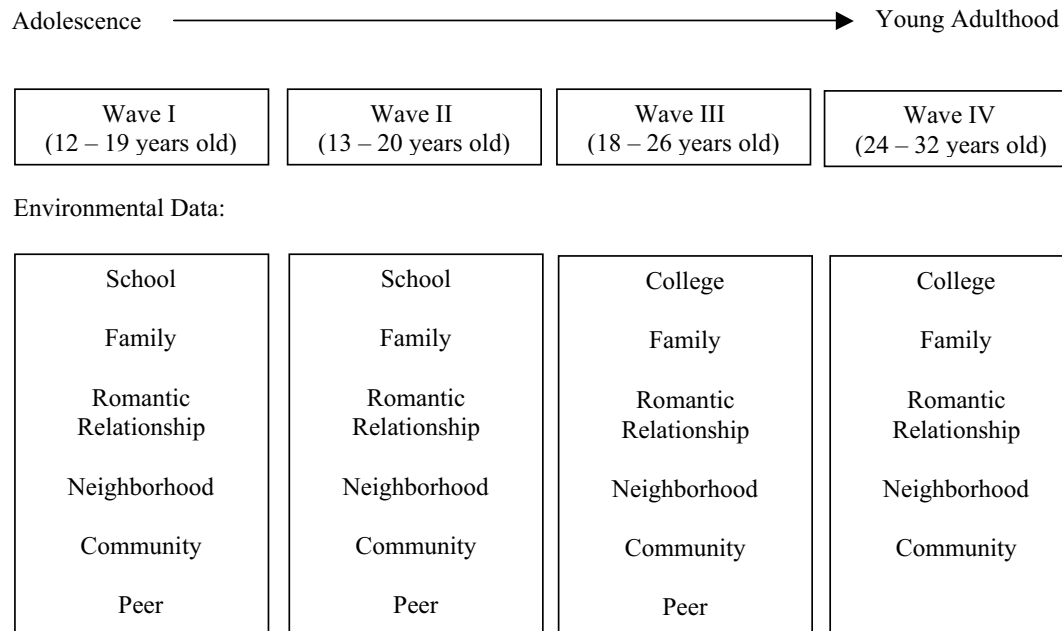
Table 3

Response Rates (%) by Type of Pairs in Genetic Sample at Wave II and Wave III

Type of pair	Pairs interviewed (Wave I)	Pairs eligible (Wave II) †	Response rate (Wave II) ‡	Pairs eligible (Wave III) †	Response rate (Wave III) ‡
Monozygotic twin	307	306	92.48	306	92.22
Dizygotic twin	452	452	94.03	451	88.91
Full-sibling	1251	1251	94.48	1251	92.01
Half-sibling	442	441	97.96	442	86.88
Unrelated	662	662	95.17	662	87.01
Total	3114	3112	94.87	3112	89.38

Note: † A pair was considered eligible if one member met the requirements for being selected into the interview at each Wave.

‡ Only one member of the pair needed to be interviewed for the pair to be included as being interviewed.

Kathleen Mullan Harris, Carolyn Tucker Halpern, Andrew Smolen, and Brett C. Haberstick

Adolescence ──────────────────────────────────────▶ Young Adulthood

| Wave I (12 – 19 years old) | Wave II (13 – 20 years old) | Wave III (18 – 26 years old) | Wave IV (24 – 32 years old) |

Environmental Data:

School	School	College	College
Family	Family	Family	Family
Romantic Relationship	Romantic Relationship	Romantic Relationship	Romantic Relationship
Neighborhood	Neighborhood	Neighborhood	Neighborhood
Community	Community	Community	Community
Peer	Peer	Peer	

Figure 1

Add Health longitudinal design.

as longitudinal data on parental support, health risk behavior, romantic and sexual relationships, and mental and physical health. At Wave IV, when the cohort will be aged 24 to 32, more attention will be devoted to career development, family formation, social and economic achievement, and early markers of future health risks or conditions. The complete set of codebooks for all survey components in all waves of interviews can be found on the Add Health web site at www.cpc.unc.edu/addhealth/codebooks.

Genotype Data

During Wave III buccal cell DNA samples were collected from a subsample of the genetic pairs by Add Health Program Project Investigator Dr. David C. Rowe of the University of Arizona. Respondents identified as full siblings or twins at Wave I and eligible for the Wave III sample (N = 3139) were asked to supply saliva specimens for DNA analysis. Compliance was high (83%), resulting in 2612 respondents with DNA samples comprising the genetic pairs subsample. Shortly before Dr Rowe's untimely death in February of 2003, the samples were transferred to the Institute for Behavioral Genetics where genotyping efforts were initiated. The authors are indebted to Dr. Rowe for his lasting contribution to this aspect of the Add Health project. The collection of DNA expanded the utility and power of this subsample substantially by allowing behavior genetic studies of adolescent health to move from variance decomposition into anonymous components, to the testing of specific hypotheses about the influence of individual genes and their expression in the context of environmental circumstances. DNA collec-

tion, in conjunction with the rich phenotypic and environmental data gathered in Add Health, creates an important research resource for future studies.

The DNA samples were genotyped for seven candidate polymorphisms and to confirm zygosity status of twins. For additional information on the field collection methods and the laboratory methods see the monograph 'Biomarkers in Wave III of the Add Health Study' posted on the Add Health website (http://www.cpc.unc.edu/projects/addhealth/files/biomark.pdf).

Confirmation of zygosity status was determined using a panel of 11 highly polymorphic, unlinked short tandem repeat (STR) markers: D1S1679, D2S1384, D3S1766, D4S1627, D6S1277, D7S1808, D8S1119, D9S301, D13S796, D15S652 and D20S481 and the sex determining locus, amelogenin (IBG-Hvar1, see http://ibgwww.colorado.edu/genotyping_lab/). The criterion used to assign monozygosity to a twin pair was 100% concordance of all genotypes at all 12 loci. A total of 34 pairs (9%) were found to have been incorrectly assigned based on questionnaire information, including 18 pairs for whom zygosity status could not be assigned from questionnaire information.

The following loci were determined for the genetic pairs subsample (N = 2612) using methods described in Anchordoquy et al. (2003), Haberstick and Smolen (2004) and Haberstick et al. (2005):

1. the dopamine transporter 40 base pair (bp) Variable Number Tandem Repeat (VNTR) in the 3' untranslated region of the gene (DAT1, Vandenbergh et al.,1992)

Table 4

Questionnaire Sections of the Wave I In-Home Interview (1995)

Section 1	General introduction
Section 2	Daily activities
Section 3	General health
Section 4	Taught in school
Section 5	Academics and education
Section 6	Physical/functional limitations
Section 7	Access to health services
Section 8	Pregnancy, AIDS, and STD risk perceptions
Section 9	Self-efficacy
Section 10	Feelings Scale
Section 11	Household roster
Section 12	Nonresident biological mother
Section 13	Nonresident biological father
Section 14	Resident mother
Section 15	Resident father
Section 16	Relations with parents
Section 17	Motivations to engage in risky behaviors
Section 18	Personality and family
Section 19	Knowledge quiz
Section 20	Friends
Section 21	Ideal romantic relationship
Section 22	Romantic relationship roster
Section 23	Liked relationship roster
Section 24	Contraception†
Section 25	Relationship information†
Section 26	Nonrelationship history†
Section 27	Motivations for birth control†
Section 28	Tobacco, alcohol, drugs†
Section 29	Delinquency Scale†
Section 30	Joint occurrences†
Section 31	Fighting and violence†
Section 32A	Male physical development†
Section 32B	Female physical development†
Section 33	Suicide†
Section 34	Parents' attitudes
Section 35	Protective factors
Section 36	Neighborhood
Section 37	Religion
Section 38	Expectations, employment, income
Section 39	Relations with siblings
Section 40	Interviewer remarks

Note: † Audio CASI.

2. the 48 bp VNTR in the third exon of the dopamine D4 receptor gene (DRD4, van Tol et al., 1992)

3. the 43 bp (not 44 bp as originally reported) addition/deletion in the 5' regulatory region of the serotonin transporter gene (5HTTLPR, Heils et al.., 1996)

4. the 30 bp VNTR in the promoter region of the monoamine oxidase A gene (MAOA, Samochowiec et al., 1999);

5. the TaqIA single nucleotide polymorphism (SNP) in the dopamine D2 receptor (DRD2, Noble et al., 1998)

6. the T→A substitution in exon 3 (rs1801272) of the cytochrome P450-2A6*2 gene (CYP2A6*2, Tyndale & Sellers, 2001)

7. the val158met SNP (rs4680) in the catechol O-methyltransferase gene (COMT, Palmateir et al., 1999).

Access to these data by external investigators is summarized below.

Research Using the Genetic Pairs in Add Health

The data produced by Add Health have been a rich source for investigators interested in issues related to health in adolescents and young adults. The Add Health website referenced above lists around 600 papers published or in press since 1997. At least 36 of these papers include the words 'twin' or 'genetic' in their titles, indicating that the research focused on the genetic pairs samples. Most of these papers have been published within the last 3 years, signifying renewed interest in the genetic pairs samples. This increased interest is likely due in large measure to the availability of genetic information that has only recently been made available to the scientific community. We briefly summarize various association studies published or in press by Add Health investigators who have utilized these data, focusing in particular on the research at the Institute for Behavioral Genetics, where Add Health DNA samples have been analyzed and stored.

Understanding how genetic and environmental factors interact to influence a trait or behavior is an important first step for more advanced molecular genetic studies. To this end, Timberlake, Rhee, et al. (2006) examined whether self-reported religiosity moderated the impact of genetic influences on smoking initiation. Controlling for gene–environment correlation, higher self-reported religiosity was found to attenuate the additive genetic effects on smoking initiation. In a separate study of current young adult smokers, Timberlake, Haberstick, et al. (2006) reported small additive genetic and shared environmental influences that did not differ as a function of gender and a large nonshared environmental contribution. Hopfer et al. (2005) examined the heritability of alcohol consumption across adolescence and young adulthood, finding an absence of shared environmental influences and heritability estimates that declined as a function of age. That pattern of results suggested the possibility that early genetic influences on alcohol consumption are different from later genetic influences and/or that new environmental influences come into play at the different ages. Results from bivariate genetic analyses indicated that different genes were contributing at later ages, as the genetic correlation betweens Waves I and III was .43, while the nonshared

Table 5

Questionnaire Sections of the Wave III In-Home Interview (2001–2002)

Section 1	Overview and demographics
Section 2	Household roster and residence history
Section 3	Parental support and relationships
Section 4	Retrospective attention deficit hyperactivity disorder
Section 5	Relationships with siblings
Section 6	Friends
Section 7	Education
Section 8	Labor market experiences and active-duty military service
Section 9	General health and diet
Section 10	Access to health services, health insurance
Section 11	Illnesses, medications, and physical disabilities
Section 12	Social psychology and mental health
Section 13	Mentoring
Section 14	Marriage/co-habitation history and attitudes
Section 15	Economics and personal future
Section 16	Sexual experiences and sexually transmitted diseases (STDs) ‡
Section 17	Compiling a Table of Relationships ‡
Section 18	Compiling a Table of Pregnancies ‡
Section 19	Relationships in detail ‡
Section 20	BEM Inventory ‡
Section 21	Propensity for risk ‡
Section 22	Completed pregnancies ‡
Section 23	Current pregnancies ‡
Section 24	Live births ‡
Section 25	Children and parenting ‡
Section 26	Delinquency and violence ‡
Section 27	Involvement with criminal justice system ‡
Section 28	Tobacco, alcohol, drugs, self-image ‡
Section 29	Mistreatment by adults ‡
Section 30	Civic participation and citizenship
Section 31	Religion and spirituality
Section 32	Gambling
Section 33	Daily activities
Section 34	Biospecimen participation
Section 35	Interviewer's report

Note: ‡ CASI.

environmental correlation was zero. The Add Health sample was used in an investigation of the gateway theory of marijuana use by Lessem et al. (2006), who reported that marijuana users were twice as likely to use illicit drugs as young adults than nonusers of marijuana. The twin and sibling subsample enabled tests of familial and environmental factors in progression of marijuana use to other drugs in later years.

Since the characterization of several functional polymorphisms in the genetic pairs subsample a number of candidate gene association studies have been conducted. Hopfer et al. (2005) examined the relationship between three dopaminergic genes and alcohol consumption across each of the three Add Health interviews. They reported significant associations between the 9-repeat (9R) allele of the DAT1 VNTR, and TaqIA SNP of the DRD2 gene with drinking in adolescents. Interestingly, those associations were not identified at each Wave, but only at Waves I and III, respectively, raising the potential that the 9R allele could be a risk factor for early drinking problems. Timberlake, Haberstick, et al. (2006) reported that the 9R allele of the DAT1 was negatively correlated with a quantitative measure of smoking frequency and quantity, suggesting that this allele has a protective effect that reduces the likelihood that these individuals will become chronic smokers.

Guo et al. (2006) reported a significant association between the DRD4 7 repeat allele and lower body mass index scores among Hispanic and Black participants, but not White, underscoring the importance of gene–gene and gene–environment interactions.

The role of measured genotype in the relationship between childhood maltreatment and conduct problems at older ages was explored by Haberstick et al. (2005). In agreement with the original report by Caspi et al. (2002), the authors reported that children who experienced early childhood maltreatment were at a higher risk for conduct problems during young adulthood. However, they were unable to find a significant association with MAOA genotype. Two subsequent studies using independent populations have reported similar conclusions (Huizinga et al., 2006; Young et al., 2006), although these do contrast with findings from three other studies (Foley et al., 2004; Nilsson et al., 2005; Windom & Brzustowicz, 2006). Guo and Tong (in press) investigated the association between age at first sex and the 48-bp repeat polymorphism in the dopamine receptor D4 gene (DRD4) by race and ethnicity. Their findings indicate that individuals with an any-3R genotype experienced a younger age at first sexual intercourse than those with an other/other (or any-4R) genotype among Asian, White and Hispanic youth, controlling for multiple socioeconomic factors. The age at first sex did not differ between the two genotypes in the black sample. Halpern et al. (in press) examined the combined contributions of attendance at religious services and three genetic polymorphisms (DRD4, DRD2, 5HTT) to sensation seeking, a personality construct related to sexual behavior, and the number of vaginal sex partners. Contrary to hypothesis, presence of the A1 DRD2 allele was associated with having had fewer sex partners. Associations between the 5HTT allele and sex partners varied by religious attendance, but patterns of associations were also contrary to hypothesized relationships and were small in magnitude, underscoring the necessity of using more comprehensive multiple gene–multiple life experience approaches to investigations of complex behaviors.

Wave IV data collection is scheduled for 2007 to 2008. We will collect DNA from the entire Add Health sample (approximately 17,000 individuals) and will increase the overall genotyping effort substantially by covering more genes and loci within those genes using a combination of STR, VNTR and SNP assessments. The combination of longitudinal phenotypic information, genetic data, twin, family and national probability sample designs inherent in the Add Health population will enable examination of more complex predictive models and should provide a valuable resource for many years to come.

Access to Add Health Data

Add Health became a pioneering study in data sharing by establishing a data dissemination plan that (1) stipulated that project investigators had no proprietary period for analysis prior to data release — the scientific community was given access to the data at the same time as project investigators; (2) established a data security plan at the time of proposal submission to NIH; and (3) developed data sharing guidelines, policies, and procedures before the data were ready for dissemination. Add Health data sets have been distributed to researchers nationwide since the release of Wave I data in 1997 and currently there are over 3000 Add Health researchers using the released data.

Add Health releases all questionnaire data obtained in the multiple survey components across all waves of interview, and releases assay and test results on biospecimens, including STD test results and the genotype data described above. Due to the sensitive and confidential content, Add Health data sets are distributed according to a tiered data disclosure plan designed to protect the data from the risk of direct and indirect disclosure of respondent identity. The tiered data disclosure plan consists of four versions of Add Health data that differ in the amount of detail in confidential information included.

Add Health releases a (1) public-use data set that is available under minimal-use restrictions and consists of one-half of the respondents in the core sample at Wave I, chosen at random, and their longitudinal data at Waves II and III. For researchers wanting to analyze the full Add Health sample with the full complement of supplemental samples, Add Health releases (2) restricted-use contractual data. Researchers and their institutions enter into a contractual agreement with Add Health and the University of North Carolina and agree to take comprehensive precautions to protect the data from nonauthorized use. Such precautions include copying the original data set only once and storing the original CD ROM in a locked drawer or file cabinet; including a unique security variable on each data set to track unauthorized use; collapsing variable categories that have small cell sizes; saving the computer programs used to construct analysis data files, but not the data files themselves; retrieving paper printouts immediately upon output and shredding printouts no longer in use; password protecting Add Health data; signing pledges of confidentiality; and agreeing to use the data solely for statistical reporting and analysis.

To be eligible to enter into a contract, a researcher must be affiliated with a research institution or organization, have an institutional review board-approved plan for handling and storing sensitive data, and sign a data-use contract agreeing to keep the data confidential. A nonrefundable fee is also required. The fee covers administrative handling charges and entitles the receiving institution to consultation with Add Health Staff. Fees also pay for travel by Add Health staff to confirm compliance with the data-use contract. A copy of the Application for Obtaining Sensitive Data from Add Health, which includes the contract, can be downloaded from the study website: http://www.cpc.unc.edu/projects/addhealth/data/contract

Add Health also makes available (3) high security restricted data for some data sets that have special heightened security requirements. These special data sets include romantic nomination links for Wave I and II data, partner data files from Wave III, and HIV assay results data from Wave III. This time-limited 1-year, single-user contract requires additional storage and use requirements such as storing and analyzing Add Health data on a stand-alone computer unlinked to any other network and a site visit to confirm adequate security provisions are in place. Finally, Add Health does not release geographic codes (geocodes), but does provide access to a (4) Secure Data Facility at the Carolina Population Center for approved researchers to link Add Health data to other contextual data sets that are geographically defined. Researchers are asked to pay an administrative fee to use the facility and must agree that all data appended to Add Health data will become part of the Add Health data sets that are distributed to the research community.

Researchers interested in obtaining archived biospecimen samples from Add Health, including urine and buccal cell DNA, to conduct additional biospecimen analysis must submit an Ancillary Study proposal to the Add Health Principal Investigator. A review of the Ancillary Study proposal is conducted by the Biology Core in consultation with the Administrative Core of Add Health. A decision to release archived samples is based on the scientific merit of the proposed project, contribution to Add Health, and quantity of specimen requested. Researchers planning a grant submission to fund Add Health specimen analysis must include documentation of Add Health approval of specimen use in their grant application. The investigator must cover all costs associated with the provision of supplemental data to Add Health, including all Add Health staff time, preparation and shipping of biospecimens, programming, reviewing data for deductive

disclosure risks, and data transfer through the Add Health security system. In addition, the investigator must agree to the Add Health dissemination policy of no proprietary period for analysis of new data produced and paid for by the investigator.

References

Anchordoquy, H. C., McGeary, C., Liu, L., Krauter, K. S., & Smolen, A. (2003). Genotyping of three candidate genes following whole genome preamplification of DNA collected from buccal cells. *Behavior Genetics, 33,* 73–78.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297,* 851–854.

Foley, D. L, Eaves, L. J., Wormley, B., Silberg, J. L., Maes, H. H., Kuhn, J., & Riley B (2004). Childhood adversity, monoamine oxidase A genotype, and risk for conduct disorder. *Archives of General Psychiatry, 61,* 738–744.

Guo, G., North, K. E., & Choi, S. (2006). DRD4 gene variant associated with body mass: The National Longitudinal Study of Adolescent Health. *Human Mutation, 27,* 236–241.

Guo, G., & Tong, Y. (in press). Age at first sexual intercourse, genes, and social and demographic context: Evidence from twins and the dopamine D4 receptor gene. *Demography.*

Haberstick, B. C., Lessem, J. M., Hopfer, C. J., Smolen, A., Ehringer, M. A., Timberlake, D., & Hewitt, J. K. (2005). MAOA genotype and antisocial behaviors in the presence of childhood and adolescent maltreatment. *American Journal of Medical Genetics, 135B,* 59–64.

Haberstick, B. C., & Smolen, A. (2004). Genotyping of three single nucleotide polymorphisms following whole genome preamplification of DNA collected from buccal cells. *Behavior Genetics, 34,* 541–547.

Halpern, C. T., Kaestle, C. E., Guo, G., & Halfors, D. D. (in press). Gene–Environment contributions to young adult sexual partnering. *Archives of Sexual Behavior.*

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., & Lesch K. P. (1996) Allelic variation of the human serotonin transporter gene expression. *Journal of Neurochemistry, 66,* 2621-–2624.

Hopfer, C. J., Timberlake, D., Haberstick, B. C., Lessem, J. M., Ehringer, M. A., Smolen, A., & Hewitt, J. K. (2005). Genetic influences on quantity of alcohol consumed by adolescents and young adults. *Drug and Alcohol Dependence, 78,* 187–193.

Huizinga, D., Haberstick, B. C., Smolen, A., Menard, S. Young, S. E., Corley, R., Stallings, M. C., Grotpeter, J., & Hewitt, J. K. (2006). Childhood maltreatment, subsequent antisocial behavior, and the role of MAOA genotype. *Biological Psychiatry, 60,* 677–683.

Lessem, J. M., Hopfer, C. J., Haberstick, B. C., Timberlake, D., Ehringer, M. A., Smolen, A., & Hewitt, J. K. (2006). Relationship between adolescent marijuana use and young adult illicit drug use. *Behavior Genetics, 36,* 498–506.

Nilsson, K. W., Sjoberg, R. L., Damberg, M., Leppert, J., Ohrvik, J., Alm, P. O., Lindrstrom, L., & Oreland, L. (2005). Role of Monoamine Oxidase A genotype and psychosocial factors in male adolescent criminal activity. *Biological Psychiatry, 59,* 121–127.

Noble, E. P., Ozkaragoz, T. Z., Ritchie, T. L., Zhang, X., Belin, T. R., & Sparks, R. S. (1998). D2 and D4 dopamine receptor polymorphisms and personality. *American Journal of Medical Genetics, 81,* 257–267.

Palmatier, M. A., Kang, A. M., & Kidd, K. K. (1999). Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. *Biological Psychiatry, 46,* 557–567.

Samochowiec, J., Lesch, K. P., Rottmann, M., Smolka, M., Syagailo ,Y. V., Okladnova, O., Rommelspacher, H., Winterer, G., Schmidt, L. G., & Sander, T. (1999). Association of a regulatory polymorphism in the promoter region of the monoamine oxidase A gene with antisocial alcoholism. *Psychiatry Research, 86,* 67–72.

Timberlake, D. S., Haberstick, B. C., Lessem, J. M., Smolen, A., Ehringer, M. A., & Hewitt, J. K. (2006). Association between the DAT1 9-repeat allele and reduced smoking in young adults from the National Longitudinal Study of Adolescent Health. *Health Psychology, 25,* 190–197.

Timberlake, D. S., Rhee, S. H., Haberstick, B. C., Hopfer, C., Ehringer, M., Lessem, J. M., Smolen, A., & Hewitt, J. K. (2006). The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. *Nicotine and Tobacco Research, 8,* 123–133.

Tyndale, R. F., & Sellers, E. M. (2001). Variable CYP2A6-mediated nicotine metabolism alters smoking behaviour and risk. *Drug Metabolism and Disposition, 29,* 548–552.

Vandenbergh, D. J., Perisco, A. M., Hawkins, A. L., Griffin, C. A., Li, X., Jabs, E. W., & Uhl, G. R. (1992). Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics, 14,* 1104–1106.

Van Tol, H. H., Wu, C. M., Guan, H. C., Ohara, K., Bunzow, J. R., Civelli, O., Kennedy, J., Seeman, P., Niznik H. B., & Jovanovic, V. (1992). Multiple dopamine D4 receptor variants in the human population. *Nature, 358,* 149–152.

Windom, C. A., & Brzustowicz, L. M. (2006). MAOA and the 'cycle of violence:' Childhood abuse and neglect, MAOA genotype, and risk for violent and antisocial behavior. *Biological Psychiatry, 60,* 684–689.

Young, S. E., Smolen, A., Hewitt, J. K., Haberstick, B. C., Stallings, M., C., Corley, R., & Crowley, T. J. (2006). Interaction between MAO-A genotype and maltreatment in the risk for conduct disorder: Failure to replicate. *American Journal of Psychiatry, 63,* 1019–1025.