

The National Longitudinal Study of Adolescent Health (Add Health) Sibling Pairs Data

Kathleen Mullan Harris,^{1,2} Carolyn Tucker Halpern,^{1,3} Brett C. Haberstick,⁴ and Andrew Smolen⁴

¹Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

²Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³Department of Maternal and Child Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁴Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, USA

This article describes the design and phenotype and genotype data available for sibling pairs with varying genetic relatedness in the National Longitudinal Study of Adolescent Health (Add Health). Add Health is a nationally representative longitudinal study of over 20,000 adolescents in the United States in 1994–1995 who have been followed for 15 years into adulthood. The Add Health design included oversamples of more than 3,000 pairs of individuals with varying genetic resemblance, ranging from monozygotic twins, dizygotic twins, full siblings, half siblings, and unrelated siblings who were raised in the same household. Add Health sibling pairs are therefore nationally representative and followed longitudinally from early adolescence into adulthood with four in-home interviews during the period 1994–2009. Add Health has collected rich longitudinal social, behavioral, environmental, and biological data, as well as buccal cell DNA from all sample members, including sibling pairs. Add Health has an enlightened dissemination policy and to date has released phenotype and genotype data to more than 10,000 researchers in the scientific community.

■ **Keywords:** Add Health, sibling pairs, genetic data, adolescence

The National Longitudinal Study of Adolescent Health, or Add Health, is an ongoing longitudinal study of a nationally representative sample of more than 20,000 adolescents in Grades 7–12 in the United States in 1994–1995 who have been followed through adolescence and their transition to adulthood with four in-home interviews in 1995, 1996, 2001–2002, and most recently in 2008–2009 when they were aged 24–32. Embedded within the design of Add Health were oversamples of about 3,000 pairs of individuals with varying genetic relatedness, including monozygotic (MZ) and dizygotic (DZ) twins, full siblings, half siblings, and adolescents with no biological relationship but who were raised in the same household. Because all design features of the Add Health Study (described below) relate to the sibling pairs sub-sample, Add Health sibling pairs are unique in that they are nationally representative, racially and ethnically diverse, and have comprehensive social, environmental, behavioral, and biological longitudinal data from early adolescence into adulthood.

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 nation-wide team of multi-disciplinary 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.

RECEIVED 1 September 2012; ACCEPTED 1 November 2012. First published online 12 December 2012.

ADDRESS FOR CORRESPONDENCE: Kathleen Mullan Harris, University of North Carolina at Chapel Hill, Carolina Population Center, CB# 8120 University Square, 123 W. Franklin Street, Chapel Hill, NC 27516-3997, USA. E-mail: kathie_harris@unc.edu

The Add Health cohort was then followed through their transition to adulthood, and research turned to understanding the determinants and consequences of developmental and health trajectories from adolescence into adulthood. Across all interview waves, comprehensive longitudinal data on health and health-related behavior were collected, including life histories of physical activity, substance use, sexual behavior, delinquency and violence, social and romantic relationships, cohabitation, marriage, childbearing, civic engagement, education, and multiple indicators of health status based on self-report (e.g., general health, mental health, chronic illness), direct measurement (e.g., overweight status and obesity), and biological measures (e.g., from blood spots, urine, and saliva).

The study has been funded by three Program Projects over the period 1994–2013 by the National Institutes of Child Health and Human Development (P01 HD031921) with co-funding from 23 other Federal agencies and national foundations. Add Health has become a national data resource for over 10,000 Add Health researchers who have obtained more than 500 independently funded research grants and have produced thousands of peer-reviewed research articles published in multiple disciplinary journals and research outlets. Below we describe the design of Add Health and its sibling pairs sub-sample, followed by the types of data available, a summary of research findings, and data access.

Add Health Design

Add Health used a school-based design that selected 80 high schools and a paired feeder school from a stratified list of all high schools in the United States in 1994. Schools were stratified by region, urbanicity, school type (public, private, parochial), ethnic mix, and size; 79% of the schools contacted agreed to participate in the study. An in-school questionnaire was administered to more than 90,000 students in Grades 7–12 who attended these schools during the 1994–1995 school year, and school administrators also filled out a questionnaire about the school. The in-school student questionnaire provided data on the school context, friendship networks, school activities, future expectations, and a variety of health conditions. An additional purpose of the in-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 sibling pairs according to their genetic relatedness 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 gender- and grade-stratified core sample of 200 adolescents was selected from each school pair for a 90-minute in-home interview as part of the Wave I inter-

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	1,251
Half sibling	442
Unrelated sibling	662

view. The core in-home sample is essentially self-weighting, and provides a nationally representative sample of 12,105 American adolescents in Grades 7–12. 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 7–12 in the United States in 1995. However, to increase the potential number of twins for genetic analysis, twins and other sibling pairs were oversampled for the in-home sample, based on responses from the in-school survey.

Sibling Pairs Recruitment

From answers provided in the in-school survey, Add Health drew supplemental samples based on the genetic relatedness of siblings in a household. 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 Wave I sample. Full siblings occur naturally in large numbers in the core in-home sample, but half siblings and unrelated adolescents (e.g., stepsiblings, foster and adopted children, adolescents in group homes) who participated in the in-school survey (i.e., were in Grades 7–12 in 1994–1995) and lived in the same household were also oversampled. Table 1 shows the number of pairs of adolescents in the sibling pairs sub-sample who were interviewed in home at Wave I. This sub-sample includes more than 3,000 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 sibling pairs design in Add Health and the data available for sibling pairs are unprecedented for a US study of this magnitude. In all follow-up interviews, high priority has been placed on locating and re-interviewing sibling pairs to maintain the integrity of this sub-sample for longitudinal research purposes.

In addition to the sibling pairs sample, supplemental samples were also drawn based on ethnicity (Cuban, Puerto Rican, and Chinese) and physical disability. Add Health also oversampled African American adolescents with highly educated parents to provide sufficient cell sizes for analyses broken down by race and socio-economic status. Finally,

TABLE 2**Sample Sizes and Response Rates (%) for Sibling Pairs by Type of Genetic Relatedness and Wave of Data Collection**

Sibling pair type	Wave I 1995	Wave II 1996	Wave II response rate [†]	Wave III 2001– 2002	Wave III response rate [†]	Wave IV 2008– 2009	Wave IV response rate [†]
MZ twins	307	283	92.8	273	89.5	282	92.5
DZ twins	452	425	94.0	401	89.1	418	92.7
Full siblings	1,251	1,182	94.6	1,151	92.4	1,173	93.9
Half siblings	442	432	98.0	384	86.9	399	90.3
Unrelated	662	630	95.2	576	87.4	611	92.3
Total	3,114	2,976	94.9	2,804	90.1	2,883	92.7

Note: [†]Response rates are based on the following denominator and numerator: a pair was considered eligible if one member met the requirements for interview selection (note this eligibility changes over waves due to death or overseas location); only one member of the pair needed to be interviewed for the pair to be included in response category.

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. A parent, usually the resident mother, also completed a 30-minute interviewer-assisted interview at Wave I. The core sample plus the special samples produced a sample size of 20,745 adolescents interviewed in the home at Wave I with a response rate of 78.9%. The Wave I in-home sample represents the national cohort that is followed prospectively through time, and thus this innovative design remains a major strength of the longitudinal data as well.

Longitudinal Design

The Wave I in-home adolescent cohort has been followed up with three subsequent in-home interviews spanning 15 years. In 1996, all adolescents in Grades 7 through 11 in Wave I and 12th graders who were part of the sibling pairs sub-sample were re-interviewed for the Wave II in-home interview ($N = 14,738$, 88.6% response rate), thereby maintaining the integrity and size of the sibling pairs sample. In addition, a follow-up school administrator interview was conducted to measure change in school context from 1995 to 1996. The original Add Health cohort was followed through their transition to adulthood with a Wave III in-home interview in 2001–2002 when the sample was aged 18–26 years, and 15,197 original respondents from Wave I were re-interviewed with a 77.4% response rate. In addition, a sample of 1,507 partners of original respondents was also interviewed, filling quota samples of 500 married, 500 cohabiting, and 500 dating partners. Wave IV re-interviewed the original Add Health cohort as they settled into adulthood in 2008–2009 when the cohort was aged 24–32 years

($N = 15,701$) with a response rate of 80.3%. For more details on the Add Health design see Harris (2010, 2011).

Response rates for the sibling pairs sub-sample have been higher than for the overall Add Health cohort at each wave because sibling pairs are easier to locate (i.e., with more family contacts) and Add Health has placed priority on maintaining this sample for longitudinal research. Table 2 shows the number of completed in-home interviews for sibling pairs according to their genetic relatedness and their response rates across waves. Response rates are based on individuals in sibling pairs who were interviewed at Wave I and who were eligible for inclusion in the Wave II, III, and IV samples. Response rates for the sibling pairs sample are quite high, especially at Wave II, only 1 year following Wave I, when 95% of the siblings in the pairs sample were re-interviewed. Even at Wave IV, almost 15 years after Wave I, response rates are over 90% for all types of sibling pairs, 93% for both MZ and DZ twin samples.

Phenotype Data

Survey Data

Add Health is a multidisciplinary multidimensional 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 3 shows the general topical areas covered by the survey instruments across waves. The Waves I and II questionnaires 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, survey attention shifted toward early family formation, post-secondary education, labor market activity and military service, mentoring, civic participation, while maintaining the longitudinal integrity of previous data on family, friends, achievement, romantic and sexual relationships, physical and mental health, and health behavior. At Wave IV, when the

TABLE 3
Questionnaire Content Across Waves

Waves I and II	Wave III	Wave IV
Demographic	Demographic	Demographic
Family, siblings, friends	Family, siblings, friends	Family, siblings, friends
Education, work	Education, work, military	Education, work, military (records)
Physical and mental health	Physical and mental health	Physical and mental health
Daily activities and sleep	Daily activities and sleep	Daily activities and sleep
Relationships	Relationships	Relationships
Sexual and fertility histories	Sexual, & fertility histories	Sexual, & fertility histories
Substance use	Substance use	Substance use and abuse
Delinquency and violence	Involvement with criminal justice system	Involvement w/criminal justice system
Attitudes, religion	Attitudes, religion	Work attitudes and characteristics, religion
Economics, expectations	Economics, expectations	Economics, expectations
Psychological, personality	Psychological, personality	Big 5 Personality, stressors
	Children and parenting	Children and parenting
	Civic participation	Civic participation
	Gambling	Cognitive function
	Mentoring	Psychosocial factors

Note: Bolded type indicates new survey content.

cohort was aged 24–32 and settling into adulthood, more attention was 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/projects/addhealth/codebooks.

Environmental Data

One of the innovations of the Add Health design was its ability to obtain independent and direct measures of the social environment of young people. The in-school and Waves I and II in-home interviews contain unique data that characterize the family, school, peer, relationship dyad, neighborhood, community, and state contexts in which Add Health respondents lived. 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 identification number). 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.

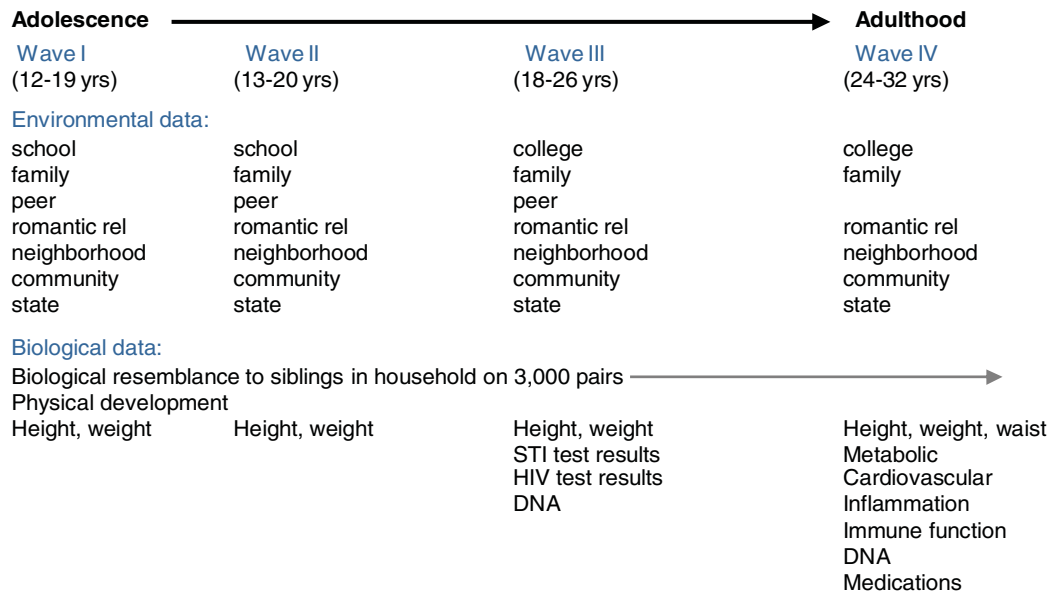
Respondents' home residences have been geocoded at each interview wave and contextual data on the neighbor-

hood, community, and state have been merged to all individual records. More than 8,500 data elements on the social and physical environment at multiple spatial levels are available across waves, including such information as race, ethnic, foreign-born, and religious denomination composition, poverty rates, crime statistics, sexually transmitted infection (STI) prevalence, divorce and child support laws, welfare policies, cigarette taxes, and the proximity and number of parks, sidewalks, recreation centers, fast food restaurants, alcohol outlets, and other physical and social characteristics of the environments in which young people live.

Figure 1 shows the developmental stages and time periods during which the Add Health cohort was followed from early adolescence into adulthood, and the types of environmental data (top panel) and biological data (bottom panel) that are available across the waves of data collection. The strength of Add Health, and thus the strength of the Add Health sibling pairs data, is the multiple levels of data that allow researchers to examine social, behavioral, environmental, biological, and genetic linkages in health and behavior across the life course from early adolescence into young adulthood.

Biological Data

The bottom panel of Figure 1 shows the biological measures available across waves. Height and weight have been measured across waves and used to track the obesity epidemic within this cohort. At Wave III, samples of urine and saliva were collected to test for STIs and human immunodeficiency virus, and buccal cell saliva was collected from the twins and full siblings in the siblings pairs subsample for DNA extraction (see Harris et al., 2006). An expanded set of biological measures was collected at Wave IV, including biomarkers of cardiovascular health (blood pressure, pulse), metabolic processes (waist circumference, HbA1c, blood glucose, lipids), immune function (EBV),

**FIGURE 1**

(Colour online) Longitudinal environmental and biological data in Add Health.

inflammation (hsCRP), and a medications log. Saliva DNA, the focus of the remainder of this article, was collected from the full sample at Wave IV, including from all sibling pairs.

Genetic Data

Add Health provides numerous opportunities for genetic research. During the Wave III (2001–2002) in-home interview, buccal cell DNA samples were collected from the twins and full sibs in the sibling pairs sample ($N \sim 2,600$) with high compliance rates (83%). These DNA samples were genotyped for seven widely studied candidate polymorphisms and the zygosity status of twins confirmed by the Institute for Behavioral Genetics at the University of Colorado Boulder. Additional information and documentation in ‘Biomarkers in Wave III of the Add Health Study’ are posted on the Add Health Web site (<http://www.cpc.unc.edu/projects/addhealth/data/guides/biomark.pdf>) and available in Harris et al. (2006).

As twin zygosity status had been based on ratings of twin resemblance during the Wave I interviews, zygosity status of the twins was again determined at Wave III using a panel of 11 highly polymorphic, unlinked short tandem repeat (STR) markers. These markers included 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.

At Wave IV (2008–2009), Add Health expanded its saliva DNA collection to include the entire sample of Add Health participants ($N = 15,701$), affording greater statistical power for genetic analyses, especially gene x environment interactions, and opportunities for replication. Saliva was collected using the Oragene collection method (Oragene, DNAgenotek, Ottawa, Ontario, Canada) and genomic DNA isolated from the Oragene solutions using ZymoResearch (Irvine, CA, USA). Consent rates for Wave IV DNA collection were high; 96% of all sibling pairs consented to provide saliva for DNA extraction, and there was little variation by type of genetic relatedness (similar to the consent rate for all Wave IV respondents, see ‘Add Health Wave IV Documentation: Candidate Genes’ at http://www.cpc.unc.edu/projects/addhealth/data/guides/DNA_documentation.pdf). The Wave IV Program Project was budgeted to genotype 10 candidate loci and a panel of short nucleotide polymorphisms (SNPs) in at least five of the following candidate genes: the dopamine transporter (DAT1, locus symbol SLC6A3); dopamine D4 receptor (DRD4); dopamine D2 receptor (DRD2); serotonin transporter (5HTT, locus symbol SLC6A4); serotonin 2A receptor (5-HT2A); Monoamine Oxidase A promoter (MAOA-uVNTR); Monoamine Oxidase A STR (MAOA[GT]n); dopamine D5 receptor (DRD5); Catechol O-methyltransferase (COMT) val158met SNP (rs4680); and insulin-like growth factor I (IGF1). Although genotype data are continually being

generated, the following polymorphisms are currently available in the full Add Health sample:

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)
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. SNP rs25531 in the Long form of the 5HTTLPR (Hu et al., 2005)
5. the 30 bp VNTR in the promoter region of the monoamine oxidase A gene (MAOA; Samochowiec et al., 1999)

Research Using the Genetic Pairs in Add Health

Genetic research in Add Health has been published in a wide range of social and biomedical science journals on topics such as substance use and dependence (e.g., Daw et al., in press; Zeiger et al., 2008), depression (e.g., Fuemmeler et al., 2009), sexual behavior (e.g., Ge et al., 2007; Halpern et al., 2007; McHale et al., 2009), political participation (e.g., Dawes & Fowler, 2009), subjective well being (e.g., De Neve, 2011), body mass index and obesity (e.g., Haberstick et al., 2010; North et al., 2010), crime and delinquency (e.g., Guo et al., 2008), education (e.g., Nielsen, 2006; Shanahan et al., 2007), suicide (e.g., Cho et al., 2006), aggression (e.g., Hart & Marmorstein, 2009), friend selection (e.g., Boardman et al., 2012; Guo, 2006), attention deficit hyperactivity disorder (e.g., Haberstick et al., 2008), conduct disorder and self-control (e.g., Schulz-Heik et al., 2010), family and peer relations (e.g., Cruz et al., 2012; Harden et al., 2008), and methodology (e.g., Medland & Neale, 2010). More than half of these publications examine the ways in which the environment interacts with genetic markers to affect health and behavior outcomes. Below we describe in more detail some illustrative examples of genetic research in Add Health.

Guo et al. (2010) examined how the dopamine transporter gene (DAT1) interacts with age (or life course stage) in relation to risk behavior (including delinquency, number of sex partners, substance use, and seatbelt use) from adolescence into young adulthood, using data on the siblings pairs from Waves I, II, and III of Add Health. They reported a protective effect of the 9R/9R genotype in the VNTR of DAT1 on risky behavior; individuals with DAT1*9R/9R compared to DAT1*Any10R, reported lower levels of risky behaviors. However, this protective effect varied according to age/life course stage, such that genetic protection is evident when the risk behavior is illegal (e.g., alcohol use and

smoking in adolescence), but vanishes when the behaviors are legal or more socially tolerated (e.g., alcohol use and smoking in adulthood). This research is important because it demonstrates how legal, as well as social, contexts can enhance or diminish genetic associations with a spectrum of risky behaviors.

Boardman et al. (2008) exploited the design of Add Health to investigate peer and school environment interactions with genetic factors associated with smoking cigarettes among adolescents. In the Add Health in-school survey, respondents nominated up to 10 of their friends who were also in-school survey participants. Adolescents receiving the most friendship nominations can be classified the 'most popular' students who shape smoking norms for the larger school community because of their social status and social connections. Boardman and colleagues assessed the smoking behavior of the most popular students and found that school norms favoring smoking (i.e., prevalence of daily smoking among the most popular students) enhanced the associations between genetic factors and daily smoking among all students. Thus, genetic contributions may not emerge unless the environment actively engages individuals in behaviors and reinforces these behaviors. Because the relative contribution of genetics to the daily use of cigarettes is conditional upon school norms related to cigarette use, there are policy opportunities to influence these norms to curb smoking behavior during the critical stage of adolescence, when initiation of smoking can set trajectories for continued use into adulthood.

Shanahan et al. (2008) examined whether the DRD2 Taq1A single nucleotide polymorphism is related to school continuation and whether social relationships compensate the DRD2 genetic risk. How much parents and their children talk about school projects, related issues, and grades captured the student-parent relationship and parental involvement in their school at Wave I (1995). School attainment was measured at Wave III (2001-2002) when Add Health participants in the sibling pairs sample were between the ages of 18 and 24 among Black and White males and females. Tests of the gene-environment interplay in this study revealed that two measures of parental involvement and the quality of school mitigated the risk of not continuing their schooling among carriers of the DRD2 Taq1A allele. These factors included a high parental socioeconomic status, high parental involvement in school, and having attended a school where a large number of students go on to attend college. These findings underscore the salience of the student-parent relationship in enhancing the likelihood of continuing education beyond high school.

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 the National Institutes of Health; 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 around the world since the release of Wave I data in 1997 and currently there are over 10,000 Add Health researchers using the released data.

Add Health releases all survey data across all waves of interview, and releases assay and test results on biospecimens, including STI test results, biomarkers, 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. Restricted data contracts require users to take comprehensive precautions to protect the data from non-authorized use and agreeing to use the data solely for statistical reporting and analysis. Data from all waves of the Add Health study are disseminated by the Inter-University Consortium for Political and Social Research as a part of their Data Sharing for Demographic Research (DSDR) project. The DSDR Add Health Web page contains the Add Health study description, publications list, documentation files, and data sets for analysis: <http://www.icpsr.umich.edu/cocoon/DSDR/STUDY/21600.xml>.

Researchers interested in obtaining archived biospecimen samples from Add Health, including urine, blood spots, and saliva DNA, to conduct additional biospecimen analysis must submit an Ancillary Study proposal to the Add Health Principal Investigator for internal and external review. 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, and must cover all costs associated with the provision of supplemental data to Add Health. 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.

In summary, Add Health provides unique longitudinal data capturing biological, psychosocial, and environmental data for nationally representative samples of sibling pairs that have been followed from early adolescence into adulthood. These data sets, and the ability to expand their coverage with ancillary studies, offer unprecedented opportunities for multi-level genetic research.

Acknowledgments

We gratefully acknowledge research support from the National Institute of Child Health and Human Development to KMH as Principal Investigator of the Add Health Program Project grant 3P01 HD31921. BCH was supported by grant AA07464. This research uses data from Add Health, a program project directed by KMH and designed by J. Richard Udry, Peter S. Bearman, and KMH at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health Web site (<http://www.cpc.unc.edu/addhealth>).

References

- Boardman, J. D., Domingue, B. W., & Fletcher, J. M. (2012). How social and genetic factors predict friendship networks. *Proceedings of the National Academy of Sciences*, *109*, 17377–17381.
- Boardman, J. D., Saint Onge, J. M., Haberstick, B. C., Timberlake, D. S., & Hewitt, J. K. (2008). Do schools moderate the genetic determinants of smoking? *Behavior Genetics*, *38*, 234–246.
- Cho, H., Guo, G., Iritani, J., & Hallfors, D. D. (2006). Genetic contribution to suicidal behaviors and associated risk factors among adolescents in the U.S. *Prevention Science*, *7*, 303–311.
- Cruz, J. E., Emery, R. E., & Turkheimer, E. (2012). Peer network drinking predicts increased alcohol use from adolescence to early adulthood after controlling for genetic and share environmental selection. *Developmental Psychology*, *48*, 1390–1402.
- Daw, J., Shanahan, M., Harris, K. M., Smolen, A., Haberstick, B., & Boardman, J. D. (in press). Genetic sensitivity to peer behaviors: 5HTTLPR, smoking, and alcohol consumption. *Journal of Health and Social Behavior*.
- Dawes, C. T., & Fowler, J. H. (2009). Partisanship, voting, and the dopamine D2 receptor gene. *Journal of Politics*, *71*, 1157–1171.
- De Neve, J. E. (2011). Functional polymorphism (5-HTTLPR) in the serotonin transporter gene is associated with subjective well-being: Evidence from a US nationally representative sample. *Journal of Human Genetics*, *56*, 456–459.
- Fuemmeler, B. F., Agurs-Collins, T., McClernon, F. J., Kollins, S. H., Garrett, M. E., & Ashley-Koch, A. E. (2009). Interactions between genotype and depressive symptoms and obesity. *Behavior Genetics*, *39*, 296–305.
- Guo, G. (2006). Genetic similarity shared by best friends among adolescents. *Twin Research and Human Genetics*, *9*, 113–121.
- Guo, G., Cai, T., Guo, R., Wang, H., & Harris, K. M. (2010). The dopamine transporter gene, a spectrum of most common

- risky behaviors, and the legal status of the behaviors. *PLoS ONE*, 5, e9352.
- Guo, G., Roettger, M. E., & Cai, T. (2008). The integration of genetic propensities into social-control models of delinquency and violence among male youths. *American Sociological Review*, 73, 543–568.
- Haberstick, B. C., Lessem, J. M., McQueen, M. B., Boardman, J. D., Hopfer, C. J., Smolen, A., & Hewitt, J. K. (2010). Stable genes and changing environments: Body mass index across adolescence and young adulthood. *Behavior Genetics*, 40, 495–504.
- Halpern, C. T., Kaestle, C. E., Guo, G., & Halfors, D. D. (2007). Gene-environment contributions to young adult sexual partnering. *Archives of Sexual Behavior*, 36, 543–554.
- Harden, K. P., Hill, J. E., Turkheimer, E., & Emery, R. E. (2008). Gene-environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. *Behavior Genetics*, 38, 339–347.
- Harris, K. M. (2010). An integrative approach to health. *Demography*, 47, 1–22.
- Harris, K. M. (2011). *Design features of Add Health*. Accessed on 1st August, 2012. Retrieved from <http://www.cpc.unc.edu/projects/addhealth/data/guides/design%20paper%20WI-IV.pdf>
- Harris, K. M., Halpern, C. T., Smolen, A., & Haberstick, B. C. (2006). The National Longitudinal Study of Adolescent Health (Add Health) twin data. *Twin Research and Human Genetics*, 9, 988–997.
- Hart, D., & Marmorstein, N. R. (2009). Neighborhoods and genes and everything in between: Understanding adolescent aggression in social and biological contexts. *Development and Psychopathology*, 21, 961–973.
- 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.
- Hu, X., Oroszi, G., Chun, J., Smith, T. L., Goldman, D., & Schuckit, M. A. (2005). An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcoholism: Clinical and Experimental Research*, 29, 8–16.
- McHale, S. M., Bissell, J., & Kim, J. Y. (2009). Sibling relationship, family, and genetic factors in sibling similarity in sexual risk. *Journal of Family Psychology*, 23, 562–572.
- Medland, S. E., & Neale, M. C. (2010). An integrated phenomic approach to multivariate allelic association. *European Journal of Human Genetics*, 18, 233–239.
- Nielsen, F. (2006). Achievement and ascription in educational attainment: Genetic and environmental influences on adolescent schooling. *Social Forces*, 85, 193–216.
- North, K. E., Graff, M., Adair, L. S., Lange, E. M., Lange, L. A., Guo, G., & Gordon-Larsen, P. (2010). Genetic epidemiology of body mass index and body mass change from adolescence to young adulthood. *Obesity*, 18, 1474–1476.
- Samochowiec, J., Lesch, K. P., Rottmann, M., Smolka, M., Sygailo, Y. V., Okladnova, O., . . . 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.
- Schulz-Heik, R. J., Shee, S. H., Silvern, L. E., Haberstick, B. C., Hopfer, C., Lessem, J. M., & Hewitt, J. K. (2010). The association between conduct problems and maltreatment: Testing genetic and environmental mediation. *Behavior Genetics*, 40, 338–348.
- Shanahan, M. J., Erickson, L. D., Vaisey, S., & Smolen, A. (2007). Helping relationships and genetic propensities: A combinatoric study of DRD2, mentoring, and educational continuation. *Twin Research and Human Genetics*, 10, 285–298.
- Shanahan, M. J., Vaisey, S., Erickson, L. D., & Smolen, A. (2008). Environmental contingencies and genetic propensities: Social capital, educational continuation, and dopamine receptor gene DRD2. *American Journal of Sociology* 114, S260–S286.
- Timberlake, D. S., Haberstick, B. C., Lessem, J. M., Smolen, A., Ehringer, M. A., Hopfer, C., & 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.
- 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., . . . Jovanovic, V. (1992). Multiple dopamine D4 receptor variants in the human population. *Nature*, 358, 149–152.
- Zeiger, J. S., Haberstick, B. C., Schlaepfer, I., Collins, A. C., Corley, R. P., Crowley, T. J., . . . Ehringer, M. A. (2008). The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco. *Human Molecular Genetics*, 17, 724–734.