Age at First Sexual Intercourse, Genes, and Social & Demographic Context: Evidence from Twins and the Dopamine D4 Receptor Gene

Guang Guo
Yuying Tong

Department of Sociology and Carolina Population Center
University of North Carolina, Chapel Hill, NC 27599

* Direct correspondence to: Guang Guo, CB#3210, Department of Sociology, University of North Carolina, Chapel Hill, NC 27599-3210 (guang_guo@unc.edu).

Age at First Sexual Intercourse, Genes, and Social & Demographic Context: Evidence from Twins and the Dopamine D4 Receptor Gene

ABSTRACT

The initiation of sexual intercourse is often viewed as one salient milestone in one's life; early initiation of sexual intercourse, however, has been shown to be a risk factor for teenage pregnancy and sexually transmitted diseases. We carried out two distinct types of genetic analysis using data Add Health. The first was an exploratory non-DNA or biometrical analysis using MZ (305 pairs) and same-sex DZ (269 pairs) twins, which tells us whether there is an aggregate genetic contribution to age at first sex. Our second analysis investigated the association between age at first sex and the 48-bp repeat polymorphism in the dopamine receptor D4 gene (*DRD4)*. Our biometrical analysis shows that MZ twins synchronize their timing of first sex to a much greater extent than the same-sex DZ twins. Our analysis of the polymorphisms in *DRD4* indicates that those with an any-3R genotype experienced a risk of first sexual intercourse 23% (P=0.020), 240% (P<0.0001), 33% (P=0.003), and 55% (P=0.036) higher than those with an other/other (or any-4R) genotype in the all-ethnicities (n=2,552), Asian, white, and Hispanic samples, respectively. The risk of first sex does not differ between the two categories of genotypes in the African American sample. These results hold in analyses when demographic and socioeconomic covariates such as gender, parental education, family structure, and community poverty were added to the regression model. Evidence from both biometrical and genetic variant analyses points to a role of genes in the timing of first sexual intercourse.

INTRODUCTION

The initiation of sexual intercourse is often viewed as one salient milestone in one's life. Early initiation of sexual intercourse, however, has been shown to be a risk factor for teenage pregnancy (Division of STD Prevention 2000**)** and sexually transmitted diseases (STDs) (Kahn et al. 2002; Latka et al. 2001; Abma and Sorenstein 2002; Ley et al. 1991; Burk et al. 1996; Burkett et al. 1992). Adolescents, both male and female, who are younger at first sexual intercourse are less likely to have used a contraceptive method (Abma et al. 1997; Sonenstein et al. 1989; Mosher and McNally 1991).

In spite of recent declines, the rate of teenage pregnancy rate in the United States is still 3 to 10 times as high as those in other industrialized countries (The Alan Guttmacher Institute 1994; Singh and Darroch 2000). The long-standing public concerns of the social and personal consequences of teenage pregnancy and childbearing is manifested in the large number of programs created to prevent these pregnancies (Kirby 1997) and in recent legislation (e.g., the Welfare Reform Act of 1996).

Each year, about 3 million adolescents are infected by a sexually transmitted disease (STD) (Institute of Medicine 1997). About 25% of all STD patients in the United States are adolescents. Chlamydia, gonorrhea, vaginitis, and pelvic inflammatory disease all have the highest prevalent rates among adolescents and these diseases become dramatically less prevalent with increasing age (Hatcher et al. 1998). Adolescents are more susceptible to STD's than adults because they have a higher probability of having multiple sexual partners and because female adolescents are biologically more susceptible to the infections (Institute of Medicine 1997). Younger ages of first sexual intercourse have been identified as a major

predictor for human papilloma virus (HPV) infection (Kahn et al. 2002; Ley et al. 1991; Burk et al. 1996; Burkett et al. 1992), cervical dysplasia, and cervical cancer (Crane 1990; LaVecchia 1986; Mfon et al 2002).

Numerous previous studies demonstrated the association between the timing of first sexual intercourse and demographic & socioeconomic factors including gender, ethnicity, parental education, family structure, and community-level factors (Sonenstein, Ku, Lindberg et al. 1998; Mott, Fondell, Hu et al. 1996; Moore, Miller, Glei, Morrison 1995; Upchurch 1998, 1999; Billy, Brewster and Grady 1994; Brewser et at. 1993; Hogan and Kitagawa 1985). Other studies explored the connection between the initialization of sexual intercourse and biological influences such as male sex hormones and pubertal development (Udry and Billy 1987; Udry et al 1985; Udry, Talbert and Morris 1986; Newcomer and Udry 1984).

A small number of studies have investigated genetic influences on age at first sex. Dunne et al. (1997) examined age at first sexual intercourse for two age cohorts of monozygotic (MZ) twins and dizygotic (DZ) twins from the Australian Twin Registry. The younger cohort were born after the period of 1952-3 and the older cohort before it. Dunne et al. reported much higher genetic contribution to the variance in age at first intercourse in the younger cohort (72% for males and 49% for females) than in the older cohort (0% for males and 32% for females). The cohort differences were interpreted as a result of gene-by-environment interaction: In the more tolerant social environment since the 1950s, individuals' biological characteristics played a more pronounced role in a person's initiation of sexual activity.

Using mainly pairs of full siblings, half siblings, and cousins and a small number (32) of same-sex twin pairs of unknown zygosity from the National Longitudinal Study of Youth

(NLSY), Rodgers et al. (1999) identified a genetic contribution to age at first sex in the all-ethnicities sample (heritability:0.37), the white sample (0.51), and the male sample (0.54). No genetic contribution to age at first sex was found in the African American sample.

Miller et al. (1999) investigated the relationship between the polymorphisms in the dopamine receptor D1 (*DRD1*), D2 (*DRD2*), and D4 (*DRD4)* genes and age at first sexual intercourse in a sample of 414 middle-class non-Hispanic European American men and women. Their analysis demonstrated a positive association between the presence of the 2 allele of *DRD2* and age at first sex.

The purpose of the present study was to investigate genetic influences on age at first sexual intercourse among 2,552 adolescents who were visited three times in 1994-5 (Wave I), 1995-6 (Wave II), and 2002 (Wave III) by the National Longitudinal Study of Adolescent Health ([Add Health]; Harris et al. 2003). In our final analysis, we incorporated genetic influences into the standard demographic and socioeconomic model of age at first sex.

RESEARCH DESIGN

We carried out two distinct types of genetic analysis. The first was an exploratory non-DNA or biometrical analysis using MZ and same-sex DZ twins, which tells us whether there is an aggregate genetic contribution to age at first sex. Our second analysis investigated the association between age at first sex and a polymorphism in the dopamine receptor D4 gene (*DRD4)*.

Biometrical Analysis. The genetic information in twin data lies in the distinction between identical (MZ) and fraternal (DZ) twins. Identical twins developed from a single zygote (one egg fertilized by one sperm) that divides into two separate cell masses within the

first two weeks of development; they are, in essence, genetic clones. Fraternal twins developed from two separate zygotes (two eggs separately fertilized by two sperms). Fraternal twins have, on average, half of their genes in common, just like any two full siblings. Combining identical and fraternal twins enables researchers to separate genetic from environmental influences without measuring genes at the molecular level. If there is a greater similarity or synchronization in the timing of first sexual intercourse within identical twin pairs than within fraternal twin pairs, then we reason that genes must have contributed to the timing of first sex. This argument holds regardless of the number of genes involved or whether the forms of the genes involved are dominant or recessive.

The twin design makes two assumptions. First, the "equal environments" assumption requires that the environments of identical twins are no more similar than are the environments of fraternal twins. If the experiences of identical twins are more similar and thus make them more alike, genetic influences would be overestimated. More similar treatment of identical twins in certain aspects of life (e.g., being dressed alike when small and/or given similar or rhyming names), however, does not automatically discredit twin studies. In our case, what is crucial is whether the special way identical twins might be treated affects the age at first sex.

The second assumption presumes that there is little or no assortative mating; this refers to the tendency of people to marry people who are like them in height, intelligence, personality, and so forth. Assortative mating could distort estimates of genetic influences in family studies. Children of similar parents would be more likely to receive the same genes for some traits than children of more dissimilar parents. For this reason, assortative mating would

exaggerate genetic similarity for fraternal twins, but it would not affect genetic similarity for identical twins because they are 100% similar genetically, with or without parental assortative mating. In a twin study, violating this assumption would underestimate genetic influences. Violations of the equal environments assumption and the assumption of assortative mating thus have opposite effects and tend to cancel each other.

The equal environments assumption has been tested in a number of ways and seems to hold for most outcome variables (Bouchard and Propping 1993). An ingenious approach used twins who were mis-categorized by their parents (Kendler eta l. 1993; Scarr and Carter-Saltzman 1979). When parents treated identical twins as fraternal twins because of mis-categorization, the mislabeled twins were as similar in behaviors and traits as identical twins who were correctly categorized, suggesting that labeling a twin pair may have only moderate consequences for the twin design.

Analysis of Genetic Variants. With the tremendous advances in molecular genetics over the past two decades (Risch 2000), twin and other biometrical studies are no longer the only strategy for studying genetic contribution to human traits and behaviors. It becomes increasingly possible to investigate complex human traits at the level of molecular genetics. In this part of the analysis, we investigated the association of the 48-bp repeat polymorphism of the dopamine receptor D4 gene (*DRD4)* with age at first sexual intercourse among 2,552 adolescents and young adults in Add Health whose DNA data are available (Harris et al. 2003).

DRD4, which maps 11p15.5 spanning 3.4 kb, is one of the five types of dopamine receptors. A functional VNTR polymorphism has been identified in the third exon in the *DRD4* gene, the region coding for the third intracellular loop of the receptor (Van Tol et al.

1992). The genetic variant is a 16 amino acid (48 bp) repeat polymorphism, which is repeated 2 to 11 times, with the two-repeat (2R), three-repeat (3R), four-repeat (4R), and seven-repeat (7R) accounting for about 97% of the alleles (Van Tol et al. 1992; Lichter et al. 1993). Extensive *in vitro* biochemical studies of DRD4 protein variants suggest that the exon III 2R and 7R alleles have decreased affinity for dopamine and transmits weaker intracellular signals in comparison to the dominant 4R protein (Asghari et al, 1995; Jovanovic et al. 1999; Oak et al. 2000). This weaker response of the 2R and 7R alleles to dopamine relative to the dominant or by far the most frequent 4R allele was the basis of the hypothesized *reward deficiency syndrome* (Blum et al 1996). It was suggested that the inhibitory neurons utilizing these "suboptimal" *DRD4* receptors would need increased dopamine for "normal" function (Swanson et al. 2000). Such increased dopamine levels were hypothesized to result from a variety of compulsive, impulsive, addictive behaviors including novelty seeking, ADHD, polysubstance abuse, alcoholism, smoking, binge eating, and compulsive gambling (Blum et al 1996).

Recently, Ding, Wang, Grady, and colleagues (Ding et al. 2002; Grady, et al. 2003; Wang et al. 2004) proposed a model of origin for the observed *DRD4*, based on DNA sequences and their nucleotide variations from worldwide representative 600 *DRD4* alleles. They showed that the minor alleles including 2R and 3R could have arisen by a one-step recombination/mutation event from the most common and ancient (>300,000 years old) 4R alleles. The behavioral traits associated with the 2R, 3R, and 7R alleles could have once given a selective advantage and produced fairly strong positive selection, resulting in a permanent presence of the these alleles in human populations. These authors suggested characterizing individuals by *DRD4* 4R versus

non-4R genotypes (including 2R, 3R, and 7R) in studies that attempt to associate *DRD4* with human behaviors and traits.

Genetic versus Demographic & Socioeconomic Influences. The advances in molecular genetics over the past two decades have identified more than 1,200 specific genes for disorders such as Huntington disease, cystic fibrosis, Duchenne muscular dystrophy, hereditary non-polyposis colon cancer, and heritable breast cancers (Botstein and Risch 2003). Almost all of these human disorders are genetically "simple" or Mendelian traits, which are basically determined by the genetic sequence at a single chromosomal locus. These simple Mendelian traits are rare. Almost all common diseases such as heart disease, hypertension, diabetes, and cancer and almost all human traits/behaviors interesting to social scientists such as age at first sexual intercourse are "complex" or non-Mendelian. Multiple genes, multiple environmental factors, and interactions among genetic and environmental factors usually influence complex traits.

Genes are unlikely to be deterministic of age at first sex itself. The more likely scenario is that a certain genetic variant is associated with a predisposition for a number of related behaviors, of which early initialization of sex is one. Genes, however, interact with environment, that is, whether the genetic predisposition leads to the behavior depends on environmental conditions.

The idea of genotype-environment interactions was well-illustrated in the recent work by Caspi et al. (2002; 2003). The 2002 article investigated the role of genotype in violent behavior among maltreated children. Boys who were maltreated early in life are at risk of becoming violent offenders. But not all children respond to maltreatment in the same way.

The study found that a functional polymorphism in the gene encoding the neurotransmitter-metabolizing enzyme monoamine oxidase A (MAOA) modifies the effect of maltreatment. Only maltreated children with a genotype generating low levels of MAOA expression tended to develop the violent behavior problem. Maltreated children with a genotype that produces high levels of MAOA activity were less affected.

In our case, the level of influence of a genetic variant on age at first sex may depend on the social and cultural meanings of the behavior for the adolescent in a particular social and cultural environment. The socioeconomic factors such as family structure, parental education, and neighborhood-level poverty (Upchurch 1998, 1999; Billy, Brewster and Grady 1993; Hogan and Kitagawa 1985) are environmental candidates for investigating such gene-environment interactions.

SAMPLES

Add Health. We used data from the National Longitudinal Study of Adolescent Health (Add Health), which started as a nationally representative sample of more than 20,000 adolescents in grades 7-12 in 1994-5 in the United States (Harris et al. 2003). The respondents have since been followed by two additional in-home interviews in 1995-6 (Wave II) and 2002 (Wave III). Add Health is school-based and the adolescents were from 134 schools. The school sample was stratified by region, ethnic mix, size, urbanicity (urban/suburban/rural), and school type (public/private/parochial).

The Twin Sample for Biometrical Analysis and the Sibling Sample for Analysis of Genetic Variants. The data for our biometrical analysis came from the twin sample within the Add Health Study, which deliberately incorporated the behavior-genetic design as a

component in an otherwise traditional survey. Our biometrical analysis used 305 pairs of monozygotic twins and 269 pairs of same-sex dizygotic twins. We excluded biological full siblings: the age difference within a pair of siblings is likely to exaggerate genetic contribution because the age difference tends to make full siblings more different than DZ twins regarding age at first sex and this difference will be attributed to genetic sources if siblings are combined with DZ twins in the analysis. Different-sex DZ twins were excluded for similar reasons. These data represent pairs of adolescents who took the exact the same questionnaire and who share the same home environment, and in most cases the same school and same neighborhood. This design creates a precious opportunity to explore the relative contributions of genetics and environment to health and health behaviors.

In Waves I and II, the classification of the twins into monozygotic and dizygotic pairs was based primarily on self-reports of confusability of appearance (Rowe et al.1999). Recently in Wave III, the zygosity of the twins were re-determined at the DNA level through a comparison of their match on 12 unlinked STR (_http://www.cpc.unc.edu/projects/addhealth/_).

In Wave III in 2002, DNA samples were collected from a subset of the Add Health sample. The subset consists of 2,597 MZ twins, DZ twins, and full biological siblings. For these individuals, genomic DNA was isolated from buccal cells using a modification of published methods (Lench et al. 1988; Meulenbelt et al. 1995; Spitz et al. 1996; Freeman et al. 1997). The specific DNA measures we used will be discussed in the Section on Measures.

MEASURES

Age at First Sexual Intercourse. At each of the three Waves, the Add Health respondents were asked about their sexual histories. They were first screened by the question: "Have you ever had sexual intercourse? When we say sexual intercourse, we mean when male inserts his penis into a female's vagina." If the respondent's answer is yes, she or he was then asked: "In what month and year did you have sexual intercourse for the very first time?" To protect confidentiality and reduce non-responses, this section of the interview was self-administered by audio-CASI (Computer Assisted Self Interview). The sensitive questions were read to respondents by means of audio headphones. Respondents were given instructions on how to complete their answers on the computer.

A number of previous studies examined the accuracy of the reported dates for age at first sex (Upchurch et al. 2002; Alexander et al. 1993; Lauritsen and Swicegood 1997; Rogers et al 1982; Siegel et al. 1998). Boys are more likely than girls to provide inconsistent reports regarding their age at first sex. The youngest boys were the least likely to provide accurate reports about their sexual experiences. Upchurch et al. (2002) showed that, at Wave II of the Add Health study, White and African American boys were more likely than White girls to revise their ages at first intercourse reported at Wave I to older ages.

The reporting errors, however, appear largely random and more importantly, they have little impact on the estimated ages at first intercourse or the effects of socioeconomic and demographic predictors of age at first sex (Upchurch et al. 2002; Lauritsen and Swicegood 1997; Wu et al. 1999). Using Wave I and II Add Health data, Upchurch et al. (2002) estimated seven sets of age at first-sex results, with each set based on a different assumption

about which reported date of first intercourse was considered true. The seven sets of results are substantively similar.

When a reporting inconsistency arose, we gave priority to the report obtained at later Wave because older adolescents tend to give more accurate reports (Siegel et al. 1998). Reporting errors may have a larger impact on our biometrical twin analysis than our analysis of genetic variants, which can be viewed for this purpose as a conventional regression analysis of age at first sex. In comparison, in the absence of DNA measures, the reported date becomes crucial. Reporting errors would reduce the within-pair synchronization or correlation regarding the timing of first sex. However, under the assumption that the rate of reporting errors does not differ by zygosity (type of twin pairs), the estimated genetic contribution (which is based on the difference between the MZ and DZ correlations) may not be affected though the baseline correlation among DZ twins will be underestimated.

DNA Preparation and Genotyping. The Wave-III Add Health collected buccal cell DNA from a subset of the Add Health sample. Genomic DNA was isolated at the Institute for Behavioral Genetics, the University of Colorado (Smolen and Hewitt, http://www.cpc.unc.edu/projects/addhealth/) using a modification of published methods (Lench et al. 1988; Meulenbelt et al. 1995; Spitz et al. 1996; Freeman et al. 1997). The average yield of DNA was 58 ± 1 μg. The best overall measurement of the quality of the DNA samples is their utility in genotype determinations. All of the Wave-III buccal DNA samples are of excellent quality and over the past 18 months have been used for the assessment of nearly 48,000 genotypes which include Short Tandem Repeats (STR), Variable Number Tandem Repeats (VNTR), and Single Nucleotide Polymorphisms (SNP).

The *DRD4* gene, which maps to 11p15.5, contains a 48 bp VNTR in the third exon (Van Tol et al. 1992). The assay used was a modification of the method of Sander et al. (Sander et al. 1997). The primer sequences (Lichter et al. 1993) were: forward, 5'-AGGACCCTCATGGCCTTG-3' (fluorescently labeled), and reverse, 5'-GCGACTACGTGGTCTACTCG-3'. This method results in PCR product of (in bp): 379 (2R), 9%; 427 (3R), 3%; 475 (4R), 65%; 523 (5R), 1%; 571 (6R), 0.8%; 619 (7R), 20%; 667 (8R), 0.9%; 715 (9R), 0.06%; and 763 (10R), 0.2%. A series of χ^2 tests each within a self-reported ethnic group (European, African American, and Hispanic) reveals no deviation from the Hardy-Weinberg equilibrium. Of the nine alleles in the polymorphism, the 2-repeat (9%), the 3-repeat (3%), the 4 repeat (65.3%), and the 7 repeat (20.4%) accounted for a total of 97.7% of the variants observed in our sample. Our analysis focused on these alleles and the genotypes associated with these alleles.

Descriptive Statistics and Demographic & Socioeconomic Predictors. Table 1 provides the descriptive statistics for the variables we used in this analysis. The 3R allele frequency (the total number of 3R alleles divided by twice the total number of individuals) in the all-ethnicities sample is about 3% with that for Asians lowest (1.6%) and whites highest (3.9%). Because only two individuals are homozygous (having two copies of the same allele) for 3R, the genotype frequency (proportion of individuals having one or two 3R alleles) is generally twice as high as the allele frequency. The genotype frequency is a more relevant statistic for our analysis than the allele frequency. The allele and genotype frequencies related to 2R and 7R alleles are not presented because they do not seem to be associated with age at first sex in our sample.

-- Table 1 about here --

Consistent with the population in the United States, the Whites are by far the largest group (57%); but Add Health over-sampled Asians (7.3%), African Americans (18.4%), and Hispanics (14.6%). A slightly higher proportion of females than males are included in the Add Health sibling sample. Family structure is a variable of four categories: two-biological families (61% in the combined sample), single-parent families (22%), step-parent families (13.2%), and other types of families (3.4%) including families with adopted and foster children. Parental education also has four categories: less than high school (11.2%), high school graduation (27.7%), some college (20.1%), and at least a college degree (36.1%). Add Health measured the level of education from both the mother and the father of a respondent. We used the higher of the two when both were available. For the 5% of the respondents missing on parental education, we created a separate category. Poverty at the neighborhood level is measured by the proportion of families living below the official poverty line in a Census block group, which is the smallest geographic area for which the Census Bureau publishes sample data. In 1990, block groups averaged 452 housing units or 1,100 individuals. The block groups are divided into low-poverty (55.6%), median-poverty (19.1%), and high-poverty (18.3%) categories. The two cutoff points for the three categories are 11.6% (the median proportion of families in a block group living below the poverty line) and 23.9% (the 75th percentile of the proportion of families in a block group living below the poverty line). We again created a separate category for those missing on neighborhood poverty (7%).

STATISTICAL PROCEDURES

We used survival analysis in both the biometrical analysis and the analysis of genetic

variants because of the right censoring in our data—a large number of individuals have not had sexual intercourse at the end of the Add Health study. Previous genetic studies on age at first sex (Dunne et al. 1997; Miller et al. 1999; and Rodgers et al. 1999) used linear regression after assuming that the age at first sex is known for all individuals in the sample. Our Add Health respondents were aged 19-26 at Wave III and the proportion that has not had first sex ranges from about 10-15% at age 26 to about 20-30% at age 19. These censored times must be dealt with in the framework of survival analysis (Cox and Oakes 1984).

Biometrical Analysis for Twin Data. For linear outcome variables, the technique of variance decomposition is routinely used to estimate the proportions of the variance in the outcome variable (e.g., body mass index) due to genetic, shared environmental, and nonshared environmental (including measurement errors) factors, respectively. The proportion of the variance due to genetic contribution is defined as heritability, which conveys the magnitude of genetic contribution to the outcome under study. The analysis can be accomplished by Pearson's correlation, structural equation models (Neale and Cardon 1992), or multilevel models (Guo and Wang 2002).

In this analysis, we estimated genetic contribution to age at first sex using twins and the shared frailty survival models or multilevel survival models (Clayton 1978):

(1) $h\ (t_{ij} \mid w_j) = w_j h_0(t_{ij}) \exp(\beta' x_{ij})$,

where the hazard function $h\ (t_{ij} \mid w_j)$ for person i in twin pair j at time t satisfies the multiplicative frailty model (1), w_j is the realized value of the random effect shared by the two members of pair j, $h_0(t_{ij})$ represents a baseline hazard, and $\beta' x_{ij}$ the usual linear predictor. The random effect W_j is assumed to have a gamma distribution with the index α,

the scale $\alpha^{-1} = \phi$, and the density $f(w_j) = w_j^{\alpha-1} e^{-\alpha w_j} \alpha^{\alpha} / \Gamma(\alpha)$, with mean $E(W_j) = 1$ and $\text{var}(W_j) = \phi$. The assumption that the mean frailty is one entails no loss of generality as long as the baseline hazard in (1) includes a constant.

In the case of linear outcome variables, genetic contribution in terms of heritability can be calculated from the random parameters (within- and between-pair variances) from multilevel or random-effects models. In the case of survival data, genetic contribution can be calculated using the between-pair variance ϕ from a gamma shared frailty model. This calculation is based on an interpretation of ϕ first suggested by Clayton (1978). He showed that the ratio of the two hazards is constant over time and equal to $1+\phi$ (2). In the ratio, the numerator is the conditional hazard of one member at age t_1 given that the other member died at age t_2; the denominator is the conditional hazard of one member at age t_1 given that the other member has survived at least to age t_2.

(2) $$\frac{h(t_1 \mid T_2 = t_2)}{h(t_1 \mid T_2 > t_2)} = 1 + \phi .$$

An estimated ϕ of 0.5, for example, would suggest that the hazard of the index member is raised by 50% if the other member in the pair died at *t*, relative to what it would be if the other member had not died by *t*. Thus, ϕ may be considered a measure of the within-pair synchronization or correlation of the age at first sex. If ϕ calculated from the MZ sample is significantly larger than that from the DZ sample, we will consider it evidence for genetic contribution to the timing of first sexual intercourse.

Analysis of Genetic Variants. This analysis consisted of two steps. First, we compared, in graphs, the life-table based probability of having had first sexual intercourse at given ages

across genotypes and ethnicities. Second, we used the following piece-wise exponential survival regression model to estimate the association between the *DRD4* genotypes and age at first sex

(3) $h\ (t_{ijk}) = h_0(t_{ijk})\exp(\beta' x_{ij})$,

Where *k* indexes the intervals within which the hazard is assumed to be constant. Model (3) could be estimated by any computer procedure that estimates the Poisson regression (Holford 1980; Laird and Olivier 1981) if the individuals in our DNA sample were independent. To control for the dependence among the siblings in our Add Health data, we estimated (3) using the generalized estimating equations (GEE) (Liang and Zeger 1986), which has long been established in the statistical literature as a standard approach for addressing correlated dadta. We have also estimated (3) using shared frailty models, which control for the correlation in the data by shared random effects (Clayton 1978); the estimates were substantively identical.

RESULTS

Table 2 presents a biometrical estimate of genetic contribution to age at first sexual intercourse and the coefficients, hazard ratios, and p-values from shared frailty survival models. These results were based on the Add Health samples of MZ and same-sex DZ twins. Only same-sex pairs were included in the DZ twin analysis in order to make them as comparable as possible to MZ twins except for the proportion of shared genetic heritage. The key estimate for our purpose is $\hat{\phi}$, the variance of the shared random effect or frailty, which was estimated to be 0.173 (P=0.004), 0.314 (P<0.001), and 0.024 (P=0.395) for the combined twin sample, the MZ twin sample, and the same-sex DZ twin sample, respectively. The estimate of 0.314 suggests that the risk of first sex is heightened by 31.4 % for the index twin

if the other twin in an identical twin pair has had sex already as compared to what it would be if the other twin has not had sex. In contrast, among the same-sex fraternal twins, the risks of first sex do not appear to be related.

-- *Table 2 about here* --

The three shared frailty survival models in Table 2 all included the standard parameters for the baseline hazards, gender, and ethnicity. Compared with White Americans, Asians experienced much lower rate of first sex (about 42% lower in the all-twins model) and African Americans had substantial higher rate of first sex (about 28% in the all-twins model). The rate of first sex peaked at ages 17-18 and then declined thereafter.

The models in Table 3 included the same set of parameters when appropriate; but only the random parameters were presented. Within both the male and female samples, the MZ twins, not the same-sex DZ twins, seem to synchronize in the timing of first sex. The random parameter estimate from the white male sample was 0.694 and highly statistically significant. The same parameter estimate in the white female sample was in the same direction, but not statistically significant.

-- *Table 3 about here* --

Figure 1 shows the proportion of those having had first sexual intercourse at given ages by gender (Panel 1) and by ethnicity (Panel 2) for the Add Health sibling sample (N=2,552). About 60% of the adolescents have had first sex by 18 and slightly higher proportions of the girls have had first sex than boys starting from the ages of 17-18. Substantial higher proportions of African American adolescents have had first sex, especially in the adolescent

years, than in the other ethnic groups. Similar proportions of White and Hispanic adolescents have had first sex and the proportions for Asians are much lower. Figure 2 is the same as Figure 1 except that the latter is based on the entire Add Health sample of 20,198 individuals. The proportions of having had first sex by gender and ethnicity in the sibling sample are very similar to that in the entire sample, indicating that our results from the sibling sample may be generalized to the entire Add Health sample.

-- Figures 1 and 2 about here --

Figure 3 provides the exploratory results for the association between the any-3R genotype in *DRD4* and the risk of first sex. Panel 1 of Figure 3 gives the proportion of those having had first sex for all ethnicities by the two genotypes: those with any 3R and those with none. Higher proportions of those with the any-3R genotype initialized sexual intercourse in adolescence than those with other genotypes, suggesting an association between the any-3R genotype and a higher risk of initializing sexual intercourse in adolescence. Panel 2 of Figure 3 is the same as Panel 1 except that the reference category (the other genotypes) only contains the any-4R genotype, which is the dominant genotype and accounts for more than 65% of all alleles in the sample. The findings in Panels 1 and 2 are almost identical.

-- Figure 3 about here --

In Panels 3-6, we present the proportion of those having had first sex at given ages by genotype for Asian, White, Hispanic, and African Americans, respectively. For Asians, the difference between the two genotypes is striking. The 3 repeat is a rare allele and only six individuals in the Asian sample belong to the any-3R genotype. Yet, by age 18-19, all with at least one 3R allele have initialized sexual intercourse. These six Asians had first sex at ages

13, 14, 15, 17, 18, and 19, respectively. In contrast, about 50% of the Asians with the none-3R genotype were still virgins at 19. In both the white and Hispanic samples, the proportions of having had first sex for the any-3R genotype are higher than those for the none-3R genotype. This pattern of a higher risk for the any-3R genotype appears to be reversed in the African American sample, in which the any-3R genotype is associated with a lower risk of having initialized first sex in adolescence. However, this apparent difference in the African American sample was not statistically significant even at the level of 10% in any of regression analyses we have conducted (to be presented).

The exploratory results suggested by Figure 3 were re-estimated in piece-wise exponential GEE survival models for the all-ethnicities sample as well as for the ethnicity-specific samples using the Add Health sibling data. The coefficients, hazard ratios, and p-values of these models are presented in Table 4. The association of the any-3R genotype with age at first sexual intercourse is evident. The all-ethnicities model suggests that the risk of initializing sexual intercourse is 23% (P=0.02) higher for those with an any-3R genotype than those with other genotypes. This increase in the risk of first sex associated with the any-3R genotype is 240% (P=<0.0001) for Asians, 33% (P=0.003) for White Americans, and 55% (P=0.036) for Hispanics. The estimated genotype effect in the African American sample is not statistically significant (P=0.12). The other parameter estimates are very similar to those presented in Table 1 based upon considerably smaller samples.

-- Table 4 about here --

In Table 5, we present models that estimate the effects of genetic variants side by side

with the effects of not only gender and ethnicity, but also family structure, parental education, and poverty at the neighborhood level. The percentage increases in the risk of first sex for the any-3R genotype in Table 5 were comparable to those in Table 4: 23% (P=0.016) for the all-ethnicities sample, 233% (P=0.0001) for Asians, 28% (P=0.012) for white Americans, and 69% (P=0.006) for Hispanics. The effect of the genetic variant was again not statistically significant for African Americans (P=0.21).

-- Table 5 about here --

In the all-ethnicities sample, growing up in a single-parent family or a step-parent family raised the risk of first sex by 23% (P=0.001) or 36% (P<0.0001), respectively, as compared with growing up in a two-biological-parents family. The effects of family structure in the black sample were very similar to those in the all-ethnicities sample whereas the effects in the Asian sample are not statistically significant. For Hispanics, growing up in a single-parent family appeared to make one particularly vulnerable, being associated with an increase of 70% (P<0.0001) in the risk of first sex. For whites, only step-families are associated with a 30% higher risk of first sex (P=0.002). In the all-ethnicities sample, parents' college degree was associated with a 26% decrease in the risk of first sex for their children, relative to those whose parents had a high school diploma. A similar effect of parental effect was found among Asians and white Americans, but not among Hispanics or African Americans. Poverty at the neighborhood level did not seem to have an impact on the risk of first sex in addition to the included individual and familial predictors.

Table 6 presents two models of age at first sex with demographic and socioeconomic predictors with one model based on the sibling sample of 2,552 individuals and the other

based on the entire Add Health sample of 20,198 individuals. The two sets of results are quite similar suggesting that our findings based on the sibling sample may be generalizeable to the US adolescent population.

-- Table 6 about here --

SUMMARY, DISCUSSION, AND CONCLUSION

Biometrical Analysis. The biometrical analyses of twins (Tables 2-3) show that the identical twins synchronized in the timing of first sexual intercourse to a substantially greater extent than fraternal twins, suggesting a significant genetic contribution to the timing of first sex. This finding was demonstrated repeatedly in four samples: the entire sample consisting of identical and same-sex fraternal twins of both genders and all ethnicities, the male sample, the female sample, and the white male sample. The finding in the white female sample was similar in magnitude, but only statistically significant at the level of 10%. This analysis was not performed for the African, Hispanic, or Asian samples because of the sharply reduced samples sizes. More specifically, our analyses suggest that the risk of having first sex for an identical twin was raised by 20-60% depending the sample of the analysis if the other twin in the pair had already had first sex, relative to the case in which the other twin was still a virgin. This synchronization was not found among fraternal twins.

For the absence of the synchronization among the DZ twins, we provide two explanations. First, no or low synchronization among DZ twins could be observed even when genetic contribution is substantial when (1) most of the shared influences are genetic rather than environmental AND (2) the behavior trait under consideration is recessive, that is, the trait is manifest only for homozygotes (the individual has two identical alleles at the locus).

(1) is consistent with one of the main findings of behavioral genetics that shared environmental influences are generally surprisingly small and frequently negligible for a wide variety of outcome variables (Boomsma et al. 2002; Plomin et al. 1997). (2) reasoned that because all MZ twins and only about a quarter of DZ twins share two alleles IBD (identical by descent), a recessive behavioral trait has a much higher probability of manifestation among MZ twins than DZ twins.

Our second and more plausible explanation lies with errors in reporting age at first sex. In an investigation of data accuracy using Waves I and II data of Add Health, Upchurch and colleagues (2002) showed that the differences in reported ages between Waves I and II averaged 4.5 months, ranging from 2.5 months for White girls to 6.9 months for African American boys. They concluded that the reporting problems were largely random and had little impact on the estimation of the effects of socioeconomic factors on age at first sex in a fixed-effects or observed-effects regression framework. Our data and analytical strategy, in contrast, are very different: we used twins in the absence of DNA measures and random-effects models to estimate genetic contribution. The reported age of first sex is a crucial piece of information for this approach. The dependence of our approach on the data accuracy is much greater than either Upchurch and colleagues' fixed-effects analysis or our own fixed-effects analysis of the *DRD4* gene.

We assume that (1)[1] the random reporting errors reduce the within-pair synchronization in age at first sex in twin analysis and (2) the reduction in the synchronization for MZ and DZ

[1] In the linear case without censoring, random reporting errors would increase the variance of the age at first sex for twin 1 and twin 2, but not affect the covariance between the two twins, and thus reduce the within-pair correlation.

twins is approximately the same. Then, a reduction similar in strength across the two types of twins would underestimate the synchronization (the parameter ϕ) in both MZ and DZ twins, but the difference in the sychcronization of the timing of first sexual intercourse between MZ and DZ twins would remain approximately the same, which is fundamental evidence for a genetic contribution to age at first sex. Thus, our twin analysis could have correctly estimated the genetic contribution even if the within-pair synchronization in age at first sex among both MZ and DZ twins is under-estimated.

Analysis of Genetic Variants. Our analysis at the molecular level showed consistently that the any-3R genotype in the dopamine D4 receptor gene (*DRD4*) was associated with a much higher risk of initializing sexual intercourse than that associated with other genotypes, which include the genotypes associated with the 4R allele, the 7R allele, and the 2R allele. We combined the one-3R and the two-3R genotypes because of the extremely small sample of the two-3R genotype (only two individuals). In separate analyses, we included in the reference category only the dominant 4R allele; our basic finding remained unaltered (data not shown). We also compared the genotypes associated with the 2R allele and the 7R allele with those associated the most frequent 4R allele; no differences were found in the timing of first sexual intercourse among these genotypes (data not shown).

The initial set of evidence was provided by Figure 3, which shows the proportion or probability of having had first sex by given ages, by genotype, and by ethnicity. Among Asians, White Americans, and Hispanic Americans, the differences by genotype took place mainly during adolescence before ages 18-20. Between ages 20 and 26, the differences actually decreased for white Americans and Hispanics. The absence of genetic contribution in

the African American sample is consistent with Rodgers et al. (1999)'s finding using the NLSY sibling data.

The heightened risk was estimated to be 240% for Asians, 33% for white Americans, and 55% for Hispanics, respectively (Table 4). Rather than viewing these dissimilar estimates across ethnicities as disparate, we found them interesting and revealing. The influence of the 3R genotype on age at first sex is likely to vary by the particular social environment the individual lived in. In our case, each ethnicity may represent a different set of environmental conditions for age at first sex. The social pressure for being sexually active among adolescents has long been thought to vary by ethnicity (DeLamater 1981; Udry and Campbell 1994). Our own analysis (Tables 4 and 5) showed that Black Americans had a higher risk (30%) of first sex than White Americans, Asians had a much lower risk (43%) than Whites, and Hispanics had a risk that is similar to that for Whites. Our analysis also showed that the levels of these ethnicity-specific risks of first sex were inversely related to the increased risk of first sex associated with the any-3R genotype. Asians had by far the lowest risk of first sex, but for Asians, the percentage increase in the risk of first sex associated with the risky genotype is by far the highest. African Americans had the highest risk of first sex, but the risky genotype did not increase the risk significantly for this sample. These preliminary findings concerning gene-environment interactions suggest the hypothesis that the genetic predisposition for initialization of sex might play a smaller role in an environment in which cultural pressure for early sex is strong than that in an environment in which such cultural pressure is milder. In the former environment, the cultural pressure may be so overwhelming that the influence from the genetic predisposition might be suppressed.

Adding the usual socioeconomic predictors of age at first sex (Table 5) does not alter the findings regarding the 3R genotype indicating that there might be little correlation between the 3R genotype and the family structure, parental education, and neighborhood poverty. We estimated models with interactions terms between the 3R genotype and the socioeconomic predictors, the resulting models did not add significant explanatory power to the models in Table 5 (data not shown). Our interaction models were likely to be under-powered because of the small proportions of the 3R genotype in our samples.

Because of the importance of the mesolimbic dopaminergic system in brain reward circuits and its implications in the reinforcing effects of many drugs and addictive behaviors, polymorphisms of genes in the dopaminergic system are plausible functional candidate genes for understanding the timing of first sexual intercourse. For *DRD4*, most of biochemical work has focused on the 2R , 4R, and 7R alleles of *DRD4* (Ding et al. 2002; Wang et al. 2002; Asghari et al. 1995) probably because of the rarity of the 3R allele (3%). Nevertheless, recent work by Ding, Wang, and colleagues has thrown some light on why the 3R genotype might be related to higher risk of first sex.

Ding, Wang, and colleagues sequenced the entire *DRD4* locus from 103 individuals homozygous for 2R, 4R, or 7R variants of the 48 bp VNTR in *DRD4* using individuals of African, European, Asian, North and South American, and Pacific Island ancestry. While the 4R/4R homozygotes display little linkage disequilibrium (LD) over the region examined, the 7R/7R individuals exhibit dramatically stronger LD surrounding the 7R allele. Based on the evidence, they propose that the 4R allele has been the most common allele throughout most of early human history and that the 7R started as a rare mutation about 40,000-50,000 years ago.

They concluded that the *DRD4* 7R allele arose recently and underwent positive selection in certain environments. They generalized this argument to the 2R allele as well as the other minor alleles including the 3R allele in *DRD4* and suggested comparing individuals with the ancestral 4R genotype versus individuals with the non-4R genotype in testing gene-behavior associations (Grady et al. 2003). As a non-4R genotype, the 3R possibly also arose more recently as compared with the 4R genotype and was selected for a reproductive advantage in a certain environment, resulting in the association with adolescent sexual behavior in contemporary US society.

We addressed the potential impact of population structure (Cardon and Palmer 2003) by stratifying our analysis by self-reported ethnicity (Tables 4 and 5). As we have already discussed, the findings vary considerably by ethnicity. The finding (the coefficient for the 3R genotype) in the all-ethnicities sample seemed to be a weighted average of the findings from the ethnicity-specific analyses. There was no evidence that population structure generated the finding in the all-ethnicities sample. We have recognized the limitation of investigating one polymorphism in the *DRD4* gene. Although it is biologically plausible that the functional 3R repeat variants are causally related to the timing of first sexual intercourse, an alternative explanation is that other functional variants within the *DRD4* gene or in an adjacent region are in linkage disequilibrium with the 3R *v*ariants. These other variants could be the real early-sex-predisposing polymorphisms.

Additional work is needed to replicate our findings that use additional genetic variants within and near the *DRD4* gene. Nevertheless, our work in both biometrical and DNA analyses has produced evidence pointing to a role of genes in the timing of first sexual

intercourse. Our work also suggests a case of gene-environment interaction; our future work will attempt to identify the specific environmental factors that were proxied by ethnicity and that might have moderated the expression of the risky genotype in the timing of first sexual intercourse.

ACKNOWLEDGEMENT

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (www.cpc.unc.edu/addhealth/contract.html).

REFERENCES

Abma J, Chandra A, Mosher W, et al. Fertility, family planning, and women's health: New data from the 1995 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 23(19). 1997.

Abma JC and Sorenstein FL. 2002. Sexual Activity and Contraceptive Practices among Teenagers in the United States, 1988 and 1995 Series 23: Data from the National Survey of Family Growth No. 21. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES.

Alexander CS, Somerfield MR, Ensminger ME, Johnson KE, and Kim YJ. 1993. Consistency of adolescents' self-report of sexual behavior in a longitudinal study. *Journal of Youth and Adolescence* 22: 455-471.

Asghari, V., S. Sanyal, et al. (1995). "Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants." J Neurochem 65(3): 1157-65.

Billy JO, Brewster KL, Grady WR. Contextual effects on the sexual behavior of adolescent women. Journal of Marriage and the Family 56(2):387–404. 1994.

Blum K, Cull JG, Braverman ER, Comings DE. Reward deficiency syndrome. *American Scientist* 1996; **84:** 132–45.

Boehnke M, Langefeld CD (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet 62:950–961.

Boomsma D, Busjahn A and Peltonen L. 2002. Classic twin studies and beyond. *Nataure Reviews Genetics* 3: 872-882.

Botstein, David and Neil Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics Supplement* 33: 228-237.

Brewster KL, Billy JOG, Grady WR. Social context and adolescent behavior: The impact of community on the transition to sexual activity. Social Forces 71(3):713–40. 1993.

Bouchard, TJ Jr. and Proppong P. (eds.). 1993. *Twins as a tool of behavioral genetics*. New York: Wiley.

Burk RD, Ho GY, Beardsley L, Lempa M, Peters M, Bierman R. Sexual behavior and partner characteristics are the predominant risk factors for genital human Papilloma virus infection in young women. J Infect Dis 1996; 174:679-89.

Burkett BJ, Peterson CM, Birch LM, Brennan C, Nuchols ML, Ward BE, et al. The relationship between contraceptives, sexual practices, and cervical human papilloma virus infection among a college population. J Clin Epidemiol 1992;45:1295-302.

Cardon LR and LJ Palmer. 2003. Population stratification and spurious allelic association. *The Lancet* 361: 598-603.

Caspi, A., J. McClay, et al. (2002). "Role of genotype in the cycle of violence in maltreated children." Science 297: 851-854.

Caspi, A, Sugden, K et al. (2003). Influence of life stress on depression: moderation in the 5-HTT gene. Science 301: 386-389.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**: 141-151.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, New York.

Crane LA. Cervical intraepithelial neoplasia: the role of age at first coitus in its etiology. J Reprod Med 1990;35: 246-59.

Curtis D (1997) Use of siblings as controls in case-control association studies. Ann Hum Genet 61:319–333.

DeLamater J. 1981. The social control of sexuality. *Annual Review of Sociology* 81: 263-290.

Division of STD Prevention. Sexually transmitted disease surveillance, 2000. Department of Health and Human Services, Atlanta: Centers for Disease Control and Prevention (CDC).

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, et al (2002): Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci U S A* 99:309 –314.

Dunne, M.P., N.G. Martin, D.J. Statham,, W.S. Slutske, S.H. Dinwiddie, K.K. Bucholz, F. Madden, & A. C. Heath. 1997. "Genetic and Environmental Contributions to Variance in Age at First Sexual Intercourse." *Psychological Science* 8: 211-216.

Freeman, B., Powell, J., Ball, D., Hill, L., Craig, I.& Plowmin, R. DNA by mail: an inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations. Behavior Genetics 1997;27:251-257.

Grady DL, Chi HC, Ding YC, Smith M, Wang E, Schuck S, Flodman P, Spence MA, Swanson JM, Moyzis RK (2003) High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. Mol Psychiatry 8:536–545.

Guo, Guang and Jianmin Wang. 2002. The Mixed or Multilevel Model for Behavior Genetic Analysis. *Behavior Genetics* 32: 37-49.

Harris, Kathleen Mullan, Francesca Florey, Joyce Tabor, Peter S. Bearman, Jo Jones, and J. Richard Udry. 2003 The National Longitudinal Study of Adolescent Health: Research Design [WWW document]. URL: http://www.cpc.unc.edu/projects/addhealth/design.

Hatcher R, Trussell J, Stewart F, et al. Contraceptive technology (17th revised edition). New York: Ardent Media, Inc. 1998.

HOGAN, D. P. & KITAGAWA, E. M. (1985) The impact of social status, family structure, and neighborhood on the fertility of black adolescents. *Am. J. Sociol.* **90,** 825–855.

Holford, T.R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics* **36**: 299-305.

Institute of Medicine. 1997. The hidden epidemic. Washington: National Academic Press.

Jovanovic V, Guan HC, Van Tol HH (1999) Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. Pharmacogenetics 9:561–568.

Kahn, JA et al. The interval between menarche and age of first sexual intercourse as a risk factor for subsequent HPV infection in adolescent and young adult women. J Pediatr 2002; 141:718-23.

Kendler KS, Neale, MC, Kessler RC, Health AC, and Eaves LJ. 1993. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* 23: 21-27.

Kirby D. No easy answers: Research findings on programs to reduce teen pregnancy. The National Campaign to Prevent Teen Pregnancy. Washington. 1997.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *J. Amer. Statist. Assoc.* **76**: 231-240.

Latka M, et al. Prevalence, incidence, and correlates of Chlamydia and Gonorrhea among young adult injection drug users. 2001. J. Substance Abuse 13: 73-88.

Lauritsen and Swicegood. 1997. The consistency of the youth risk behavior surveillance system. *Public Health Report*, 108 (Suppl. 1), 2-10.

LaVecchia C, Franceschi S, Decarli A, Fasoli M, Gentile A, Parazzini F, et al. Sexual factors, venereal diseases, and the risk of intraepithelial and invasive cervical neoplasia. Cancer 1986;58: 935-41.

Lench, N., Stanier, P. and Williamson, R. Simple non-invasive method to obtain DNA for gene analysis. Lancet 1988:1356-1358.

Ley C, Bauer H, Reingold A, Schiffman MH, Chambers JC, Tashiro CJ, et al. Determinants of genital Human Papilloma virus infection in young women. J Natl Cancer Inst 1991; 83:997-1003.

Liang, KY and Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13-22.

Lichter, J.B., Barr, C.L., Kennedy, J.L., Van Tol, H.H.M., Kidd, K.K. and Livak, K.J. A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. Hum Mol Genet 1993;2:767-773.

Meulenbelt, I., Droog, S., Trommelen, G.J., Boomsma, D.I. and Slagboom, P.E. (1995). High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. Amer J Hum Genet 1995;57:1252-1254.

Mfon S et al. Cervical cancer risk as a predictor of pap smear use in rural North Carolina. 2002. 18: 67-76.

MILLER, W. B., PASTA, D. J., MACMURRAY, J., CHIU, C., WU, H. & COMINGS, D. E. (1999) Dopamine receptors are associated with age at first sexual intercourse. *J. biosoc. Sci.* **31,** 43–54.

Moore KA, Miller BC, Glei D, Morrison DR. Adolescent sex, contraception, and childbearing: A review of recent research. Washington: Child Trends, Inc. 1995.

Mosher WD, McNally JW. Contraceptive use at first premarital intercourse: United States, 1965–1988. Fam Plann Persp 23(3):108–16. 1991.

Mott FL, Fondell MM, Hu PN, et al. The determinants of first sex by age 14 in a high-risk adolescent population. Fam Plann Persp 28(1):13–18. 1996.

Neale MC, Cardon LR, Methodology for Genetic Studies of Twins and Families (NATO ASI Series D: Behavioural and Social Sciences-Vol. 67), Kluwer Academic Publishers B.V., Dordrecht, The Netherlands, 1992.

Oak JN, Oldenhof J, Van Tol HHM (2000) The dopamine D4 receptor: one decade of research. Eur J Pharmacol 404:303–327

Plomin, Robert, John C. DeFries, Gerald E. McClean, and Michael Rutter. *Behavioral Genetics, 3rd ed, 1997* New York: W.H. Freeman and Company.

Risch, Neil J. "Searching for Genetic Determinants in the New Millennium." *Nature* (405), 2000: 847-850.

Rogers, JL, Billy JOG, and Udry JR. 1982. The rescission of behaviors: Inconsistent responses in adolescent survey data. *Social Science Research* 11: 280-296.

RODGERS, J. L., ROWE, D. C. & BUSTER,M. (1999) Nature, nurture, and first sexual intercourse in the USA: Fitting behavioural genetic models to NLSY kinship data. *J. biosoc. Sci.* **31,** 29–41.

Rowe, D., K. Jacobson and Evan van den Oord. 1999. "How People Make Their Own Environments: A Theory of Genotype -> Environmental Effects." Child Development 54: 424-435.

Sander, T., Harms, H., Dufeu, P., Kuhn, S., Rommelspacher, H. and Schmidt, L.G. Dopamine D4 receptor Exon III alleles and variation of novelty seeking in alcoholics. Amer J Med. Genet 1997;74:483-487.

Scarr S and Carter-Saltzman L. 1979. Twin method: Defense of a critical assumption. *Behavior Genetics* 9: 527-542.

Siegel, DM, Aten MJ, and Roghmann, KJ. 1998. Self-reported honesty among middle and high school students responding to a sexual behavior questionnaire. *Journal of Adolescent Health*: 23: 20-28.

Singh S, Darroch JE. Adolescent pregnancy and childbearing: Levels and trends in developing countries. Fam Plann Persp 32(1):14–23. 2000.

Smolen A and Hewitt J. http://www.cpc.unc.edu/projects/addhealth/.

Sonenstein FL, Ku L, Lindberg LD, et al. Changes in sexual behavior and condom use among teenaged males: 1988 to 1995. Am J Public Health 88(6): 956–59. 1998.

Sonenstein FL, Pleck JH, Ku LC. Sexual activity, condom use, and AIDS awareness among adolescent males. Fam Plann Persp. 21(4):152–58. 1989.

Spitz, E., Moutier, R., Reed, T., Busnel, M., C., Marchaland, C., Roubertoux, P.L. and Carlier, M. Comparative diagnoses of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. Behav Genet 1996;26:55-64.

Swanson JM, Oosterlaan J, Murias M, Schuck S, Flodman P, Spence MA, Wasdell M, Ding YC, Chi HC, Smith M, Mann M, Carlson C, Kennedy JL, Sergeant JA, Leung P, Zhang YP, Sadeh A, Chen C, Whalen CK, Babb KA, Moyzis R, Posner MI (2000) Attention deficit/hyperactivity disorder children with a 7-repeat allele of the dopamine receptor D4 gene have extreme behavior but normal performance on critical neuropsychological tests of attention. Proc Natl Acad Sci USA 97:4754–4759.

The Alan Guttmacher Institute. Sex and America's teenagers. New York: The Alan Guttmacher Institute. 1994.

UDRY, J. R. & BILLY, J. O. G. (1987) *American Sociological Review* **52,** 841–855.

UDRY, J. R. & CAMPBELL, B. C. (1994) Getting started on sexual behavior. In: *Sexuality across the Life Course.* Edited by A. Rossi. University of Chicago Press, Chicago.

UDRY, J. R. (1988) Biological predispositions and social control in adolescent sexual behavior. *Am. sociol. Rev.* **53,** 709–722.

UDRY, J. R., BILLY, J. O. G.,MORRIS,N. M., GROFF, T. R. &RAJ,M.H. (1985) Serum androgenic hormones motivate sexual behavior in adolescent human males. *Fert. Steril.* **43,** 90–94.

NEWCOMER, S. F. & UDRY, J. R. (1984) Mothers' influence on the sexual behavior of their teenage children. *J. Marriage Fam.* **46,** 477–485.

UDRY, J. R., TALBERT, L. M. & MORRIS, N. M. (1986) Biosocial foundations for adolescent female sexuality. *Demography* **23,** 217–230.

Upchuch, DM, Levy-Storms L, Sucoff CA, and Aneshensel CS. 1998. Gender and ethnic differences in the timing of first sexual intercourse. *Family Planning Perspective* 30: 121-127.

Upchurch, DM, Aneshensel CS, Sucoff CA, and Levey-Storms L. 1999. *J. of Marriage and the Family* 61: 920-933.

Upchurch, DM, Lillard LA, Aneshensel CS, and Li NF. 2002. Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. J. of Sex Research 39: 197-206.

Van Tol, H. H., J. R. Bunzow, et al. (1991). "Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine." Nature 350(6319): 610-4.

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

Wang E, Ding YC, et al. The Genetic Architecture of Selection at the Human Dopamine Receptor D4 (*DRD4*) Gene Locus. *Am. J. Hum. Genet. 74:931–944, 2004.*

Wise RA, Rompre P-P (1989) Brain dopamine and reward. Annu Rev Psychol 40:191–225.

Wu et al. 1999 Wu LL, Martin SP, and Long DA. 1999. Comparing data quality of fertility and first sexual intercouse histoires. Paper presented at the Data Quality in Longitudinal Surveys Conference, Hyattsville, MD.

Table 1. Descriptive statistics for the Add Health siblings sample

	All ethnicities		Asian		White		Hispanic		Black	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
3R (allele)	0.030		0.016		0.039		0.015		0.019	
Any-3R (genotype)	0.059	0.235	0.032	0.177	0.078	0.269	0.030	0.170	0.036	0.187
Ethnicity										
White	0.571	0.495								
Asian	0.073	0.260								
Black	0.184	0.388								
Hispanic	0.146	0.353								
Other	0.026	0.160								
Female	0.522	0.500	0.478	0.501	0.526	0.500	0.505	0.501	0.549	0.498
Family Structure										
2 bio. parents	0.613	0.487	0.747	0.436	0.673	0.469	0.629	0.484	0.370	0.483
Single parent	0.220	0.414	0.108	0.311	0.159	0.365	0.202	0.402	0.468	0.500
Step parent	0.132	0.339	0.091	0.289	0.149	0.356	0.126	0.333	0.094	0.292
Other families	0.034	0.183	0.054	0.226	0.019	0.137	0.043	0.203	0.068	0.252
Parental education										
High school	0.277	0.447	0.124	0.330	0.291	0.454	0.223	0.417	0.336	0.473
< high school	0.112	0.316	0.075	0.265	0.053	0.224	0.355	0.479	0.115	0.319
Some college	0.201	0.401	0.156	0.364	0.222	0.416	0.145	0.353	0.189	0.392
≥ college	0.361	0.480	0.575	0.496	0.404	0.491	0.204	0.404	0.283	0.451
Missing	0.049	0.216	0.070	0.256	0.030	0.171	0.073	0.260	0.077	0.266
Neighborhood poverty										
< 11.6 %	0.556	0.497	0.785	0.412	0.653	0.476	0.470	0.500	0.243	0.429
11.6%-23.9%	0.191	0.393	0.129	0.336	0.194	0.396	0.210	0.408	0.198	0.399
≥ 23.9 %	0.183	0.386	0.043	0.203	0.091	0.287	0.237	0.426	0.479	0.500
Missing	0.070	0.255	0.043	0.203	0.062	0.242	0.083	0.277	0.081	0.273
No. of individual	2,552		186		1,457		372		470	

Table 2. Biometrical estimate of genetic contribution to age at first sexual intercourse: coefficients, hazard ratios, and p-values from shared frailty survival models (Add Health twin samples)

	MZ + same-sex DZ twins			MZ twins			same-sex DZ twins		
	β	exp(β)	P-value	β	exp(β)	P-value	β	exp(β)	P-value
Intercept	-4.526	0.011	0.000	-4.709	0.009	0.000	-4.359	0.013	0.000
Male	--	--	--	--	--	--	--	--	--
Female	0.115	1.122	0.151	0.191	1.210	0.115	0.052	1.053	0.623
White	--	--	--	--	--	--	--	--	--
Asian	-0.540	0.583	0.021	-0.479	0.619	0.083	-0.768	0.464	0.192
Black	0.248	1.282	0.012	0.303	1.353	0.052	0.175	1.191	0.161
Hispanic	0.069	1.071	0.550	0.205	1.227	0.221	-0.083	0.920	0.600
Other	0.122	1.130	0.589	-0.005	0.995	0.989	0.250	1.284	0.376
<15	--	--	--	--	--	--	--	--	--
15-16	2.497	12.140	0.000	2.488	12.041	0.000	2.503	12.218	0.000
16-17	2.910	18.349	0.000	2.907	18.295	0.000	2.908	18.316	0.000
17-18	3.253	25.871	0.000	3.423	30.653	0.000	3.038	20.863	0.000
18-20	2.865	17.553	0.000	3.054	21.192	0.000	2.660	14.292	0.000
20+	2.617	13.698	0.000	3.059	21.299	0.000	2.042	7.703	0.000
No. of pairs	574			305			269		
-2Log L	802.34			403.6			386.08		
ϕ	**0.173**		**0.004**	**0.314**		**0.000**	**0.024**		**0.395**

Table 3. Biometrical estimate of genetic contribution to age at first sexual intercourse: coefficients, hazard ratios, and p-values from shared frailty survival models by gender and ethnicity (Add Health twin samples)

	Male		Female		White Male		White Female	
	MZ Twins	DZ Twins	MZ Twins	DZ Twins	MZ Twins	DZ Twins	MZ Twins	DZ Twins
				---Other parameters omitted---				
No. of pairs	154	136	151	133	90	81	80	71
ϕ **(P-value)**	**0.262 (0.032)**	**0.088 (0.250)**	**0.352 (0.003)**	**0.000 (0.49)**	**0.694 (0.001)**	**0.000 (1.000)**	**0.199 (0.099)**	**0.000 (0.50)**

Table 4. Association of the any-3-repeat genotype in the dopamine D4 receptor gene (*DRD4*) with age at first sexual intercourse: coefficients and hazard ratios of Piece-wise exponential GEE survival model by ethnicity (Add Health sibling samples)

	All ethnicities			Asian			White			Hispanic			Black		
	β	exp(β)	P-value	β	exp(β)	P-value	β	exp(β)	P-value	β	exp(β)	P-value	β	exp(β)	P-value
Intercept	-4.35	0.01	0.000	-5.18	0.01	0.000	-4.43	0.01	0.000	-4.30	0.01	0.000	-3.88	0.02	0.000
Other/Other	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Any 3R	**0.21**	**1.23**	**0.020**	**1.22**	**3.40**	**<0.0001**	**0.29**	**1.33**	**0.003**	**0.44**	**1.55**	**0.036**	**-0.42**	**0.65**	**0.123**
White	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Asian	-0.57	0.57	0.000												
Black	0.26	1.30	0.000												
Hispanic	0.07	1.07	0.253												
Other	0.31	1.37	0.014												
Male	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Female	0.12	1.12	0.010	0.03	1.03	0.849	0.16	1.17	0.003	0.13	1.14	0.229	0.01	1.01	0.908
<15															
15-16	2.51	12.26	0.000	2.92	18.61	0.000	2.49	12.05	0.000	2.39	10.92	0.000	2.60	13.47	0.000
16-17	2.77	15.88	0.000	3.02	20.47	0.000	2.79	16.32	0.000	0.63	1.88	0.000	2.85	17.34	0.000
17-18	3.16	23.50	0.000	3.55	34.66	0.000	3.18	24.15	0.000	2.99	19.98	0.000	3.23	25.25	0.000
18-20	2.62	13.74	0.000	3.21	24.82	0.000	2.57	13.07	0.000	2.57	13.08	0.000	2.68	14.54	0.000
20+	2.40	11.00	0.000	2.74	15.49	0.000	2.49	12.12	0.003	2.41	11.10	0.000	1.63	5.11	0.000
No. of Persons		2,552			186			1,457			372			470	
-2LogL	10,408			716			5,972			1,552			1,824		

Table 5. Association of the any-3-repeat genotype in the dopamine D4 receptor gene (*DRD4*) with age at first sexual intercourse: coefficients and hazard ratios of Piece-wise exponential GEE survival model by ethnicity with demographic and socioeconomic predictors (Add Health sibling samples)

	All ethnicities			Asian			White			Hispanic			Black		
	β	e^{β}	P-value	β	e^{β}	P-value	β	e^{β}	P-value	β	e^{β}	P-value	β	e^{β}	P-value
Intercept	-4.29	0.01	<.0001	-4.75	0.01	<.0001	-4.33	0.01	<.0001	-4.28	0.01	<.0001	-3.95	0.02	<.0001
Other/other	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Any 3R	**0.21**	**1.23**	**0.016**	**1.20**	**3.33**	**0.0001**	**0.25**	**1.28**	**0.012**	**0.52**	**1.69**	**0.006**	**-0.33**	**0.72**	**0.216**
White	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Asian	-0.51	0.60	<.0001												
Black	0.19	1.21	0.003												
Hispanic	0.06	1.06	0.323												
Other	0.32	1.38	0.005												
Female	0.11	1.11	0.009	0.029	1.03	0.870	0.15	1.16	0.006	0.07	1.08	0.497	0.02	1.03	0.788
Parameter estimates for the baseline hazards omitted															
Family Structure															
2 biology parents	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Single parent	0.21	1.23	0.001	0.01	1.01	0.973	0.04	1.04	0.628	0.53	1.70	<.0001	0.26	1.30	0.010
Step parent	0.31	1.36	<.0001	0.25	1.28	0.489	0.26	1.30	0.002	0.16	1.18	0.251	0.38	1.47	0.017
Other families	0.47	1.60	0.005	0.83	2.28	0.000	0.45	1.57	0.369	0.82	2.27	0.049	0.33	1.39	0.168
Parental education															
High school	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
< high school	-0.13	0.88	0.097	-0.65	0.52	0.124	0.09	1.09	0.524	-0.20	0.82	0.159	0.00	1.00	0.988
Some college	-0.05	0.95	0.314	-0.33	0.72	0.359	-0.05	0.95	0.502	0.12	1.13	0.384	-0.16	0.85	0.132
≥ college	-0.30	0.74	<.0001	-0.59	0.56	0.023	-0.37	0.69	<.0001	-0.11	0.89	0.430	-0.07	0.94	0.562
Missing	-0.61	0.54	0.001	-0.69	0.50	0.094	-0.52	0.60	0.320	-1.01	0.36	0.026	-0.59	0.55	0.029
Neighborhood poverty (Proportion of families with income below the poverty level)															
< 11.6 %	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
11.6%-23.9%	0.04	1.04	0.520	-0.08	0.92	0.807	-0.02	0.98	0.820	0.18	1.20	0.122	-0.03	0.97	0.836
≥ 23.9 %	-0.05	0.95	0.486	0.05	1.05	0.877	0.00	1.00	0.993	-0.13	0.88	0.347	-0.09	0.92	0.458
Missing	-0.06	0.94	0.433	-0.24	0.79	0.327	0.18	1.20	0.059	-0.38	0.68	0.120	-0.26	0.77	0.153
No. of persons		2,552			186			1,457			372			470	
-2LogL	10298.791			704.298			5917.65			1523.83			1810.865		

Table 6. Age at first sexual intercourse: coefficients and hazard ratios of Piece-wise exponential GEE survival models: Comparison between the sibling and the entire Add Health samples

	Sibling sample			All sample		
	β	e^{β}	P-value	β	e^{β}	P-value
Intercept	-4.266	0.014	<.0001	-4.097	0.017	<.0001
White	--	--	--	--	--	--
Asian	-0.522	0.594	<.0001	-0.389	0.678	<.0001
Black	0.191	1.210	0.003	0.130	1.139	<.0001
Hispanic	0.049	1.050	0.429	-0.085	0.919	0.001
Other	0.314	1.368	0.006	0.096	1.101	0.048
Female	0.102	1.107	0.012	0.050	1.051	0.002
Family Structure						
2 biology parents	--	--	--	--	--	--
Single parent	0.205	1.227	0.000	0.196	1.216	<.0001
Step parent	0.312	1.366	<.0001	0.286	1.331	<.0001
Other families	0.459	1.582	0.007	0.220	1.246	<.0001
Parental education						
High school	--	--	--	--	--	--
< high school	-0.129	0.879	0.103	-0.074	0.929	0.009
Some college	-0.057	0.945	0.302	-0.041	0.959	0.077
≥college	-0.301	0.740	<.0001	-0.202	0.817	<.0001
Missing	-0.595	0.552	0.002	-0.292	0.747	<.0001
Neighborhood poverty[a]						
< 11.6 %	--	--	--	--	--	--
11.6%-23.9%	0.031	1.032	0.579	0.072	1.075	0.003
≥ 23.90%	-0.050	0.952	0.471	0.044	1.045	0.092
Missing	-0.056	0.946	0.473	0.009	1.009	0.641
Baseline Hazard						
<15						
15-16	2.511	12.31	<.0001	2.514	12.35	<.0001
16-17	2.775	16.03	<.0001	2.691	14.74	<.0001
17-18	3.173	23.87	<.0001	2.952	19.14	<.0001
18-20	2.638	13.98	<.0001	2.512	12.32	<.0001
20+	2.427	11.32	<.0001	2.085	8.04	<.0001
No. of persons		2,552			20,198	
-2LogL		10304.19			75222.13	

Figure 1. Proportion of those having had first sexual intercourse at given ages: Add Health sibling samples

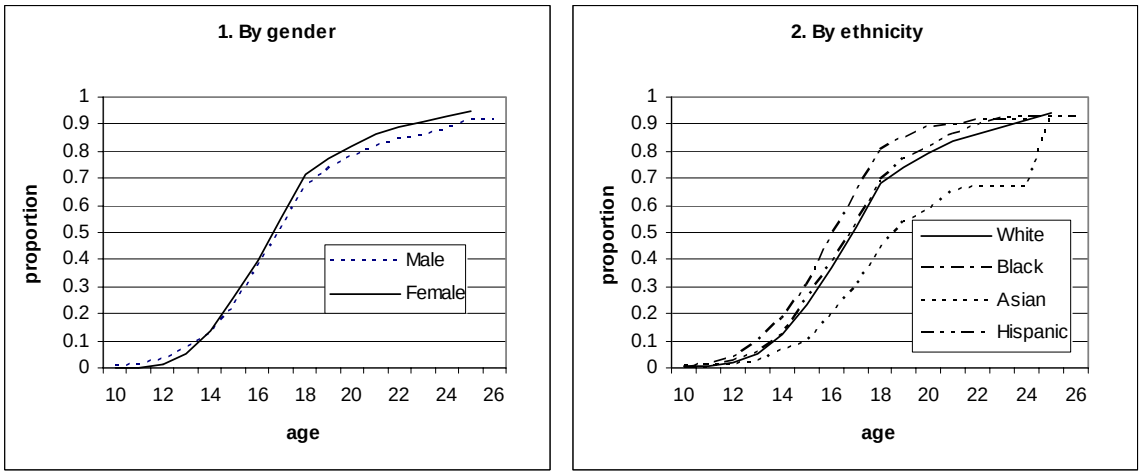

Figure 2. Proportion of those having had first sexual intercourse at given ages: Add Health entire samples

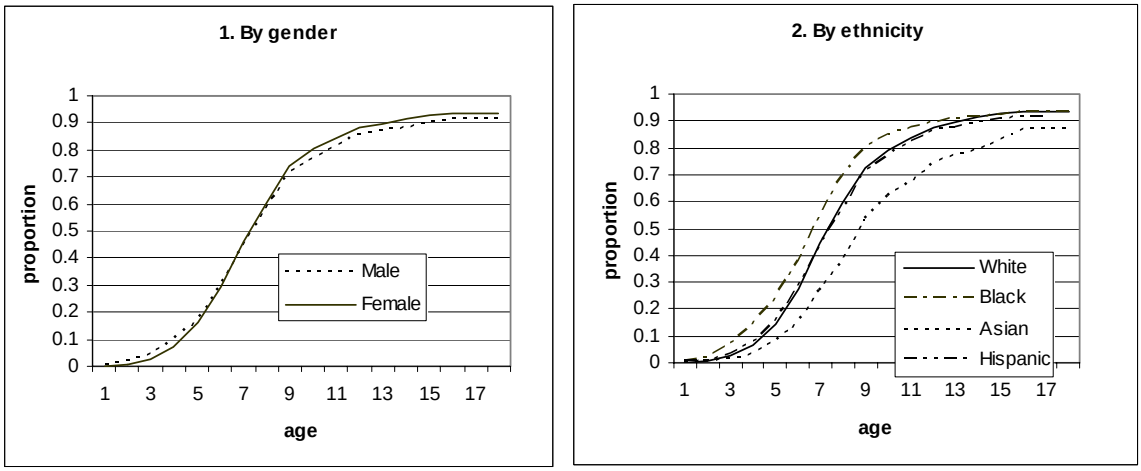

Figure 3: Proportion of those having had first sexual intercourse at given ages by genotype