GENETIC AND ENVIRONMENTAL OVERLAP BETWEEN BEHAVIORAL DISINHIBITION AND RISKY SEXUAL BEHAVIORS

by

BROOKE M. HUIBREGTSE

B.A., University of Minnesota, Twin Cities, 2010

M.A., University of Colorado Boulder, 2014

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2017

This thesis entitled:
Genetic and environmental overlap between behavioral disinhibition and risky sexual behaviors
written by Brooke Marie Huibregtse
has been approved for the Department of Psychology and Neuroscience

Michael C. Stallings

John K. Hewitt

Matthew C. Keller

Stefanie Mollborn

Soo H. Rhee

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocols #14-0527, 17-0059, & 0399.

Huibregtse, Brooke Marie (Ph.D., Department of Psychology and Neuroscience)

Genetic and Environmental Overlap between Behavioral Disinhibition and Risky Sexual Behaviors

Thesis directed by Professor Michael Stallings

Previous studies suggest that risky sexual behaviors (RSB) are highly correlated with impulsive behaviors such as substance use disorders, antisocial behavior, and novelty seeking. The comorbidity of these latter behaviors is well described by an underlying heritable factor termed behavioral disinhibition (BD). To better understand the nature of this correlation, this dissertation explores the extent to which this overlap is genetic or environmental in nature. Multivariate biometrical models with twin and adoptive samples are used in Chapters II and III to assess developmental trends in substance use behaviors and to explore the shared etiology between sex under the influence of drugs and alcohol and number of lifetime sexual partners. Chapter II included an in depth review of issues regarding measuring and defining RSB, with the goals of improving an instrument for measuring sexual behavior, improving the interpretation of several RSB variables, and selecting an optimal phenotype for use with genome wide methods. Finally, Chapter V used several genome-wide approaches to explore the genetic architecture of number of lifetime sexual partners and to test the genetic overlap with measures BD related diseases and traits (e.g. smoking, psychiatric, personality), and other fitness phenotypes.

AKNOWLEDGEMENTS

My research was supported by an institutional training grant from the National Institute on Drug Abuse (T32DA017637, John K. Hewitt).

Data for the Center on Antisocial Drug dependence was collected and supported via grants from National Institutes for Health including the National Institute on Aging (AG046938), National Institute for Child and Human Development (HD010333, HD036773), and the National Institute on Drug Abuse (DA011015, DA012845, DA021913, DA021905, DA035804 and DA015522).

Funding for data from the Minnesota Twin and Family Center is from the National Institute on Alcohol Abuse and Alcoholism (AA09376, AA11886), the National Institute on Drug Abuse (DA05147, DA13240, and DA024417) and the National Institute on Mental Health (MH0066140).

I am very grateful to Michael C. Stallings for this advice and guidance throughout graduate school. For their support and training, I would like to acknowledge my committee members, as well as the faculty and staff of the Institute for Behavioral Genetics. Thank you my fellow graduate students, from whom I have learned a great deal. A final thank you is directed toward my family, friends, and to Shaw L. Ketels—my favorite collaborator in research and in life.

CONTENTS

CHAPTER

TABLES

FIGURES

CHAPTER I

Introduction

The comorbidity of impulsive behaviors such as substance use disorders, antisocial behavior, and novelty seeking is well described by an underlying heritable factor often termed behavioral disinhibition. Though risky sexual behavior is largely explained by the same underlying vulnerabilities (namely, impulsivity; Dir, Coskunpinar & Cyders, 2014), there is no current consensus on whether risky sexual behavior should be considered a central indicator of this factor. To be considered as such, risky sexual behavior should be 1) strongly correlated with other behaviors and 2) correlated mainly due to common genetic influences.

Our overall goal is to demonstrate the shared etiology (i.e. genetic and environmental overlap) of risky sexual behavior with other indicators of behavioral disinhibition using a variety of developmental & multivariate biometric models, as well as quantitative genome wide methods.

Literature Review

Behavioral Disinhibition

Behavioral disinhibition (BD) can be thought of as "lack of constraint, tendency toward impulsivity, or inability to inhibit socially undesirable or otherwise restricted actions" (Iacono et al., 2008). Similarly, BD has been described as predisposition for high novelty seeking, impulsivity, and lack of constraint (Sher & Trull, 1994). Before epidemiological research confirmed that BD behaviors were highly comorbid, a number of researchers observed and described the frequency and extent for impulsive and restricted behaviors to cluster together. Problem behavior theory (Jessor & Jessor, 1977) posited that those behaviors which are problematic, a source of concern, or which violate social or legal norms (e.g. substance use,

delinquent behavior, early sexual intercourse, or risky driving behavior) occur together in the way that more conventional behaviors also occur together. Early factor analysis of childhood psychopathology suggested a strong relationship between "undercontrolled" behaviors (Achenbach & Edelbrock, 1978). Drawing upon animal literature, Gornstein and Newman (1980) theorized that a general disinhibitory personality with biological underpinnings linked together psychological constructs previously thought of as distinct such as "syndromes including psychopathy, hyperactivity, hysteria, antisocial and impulsive personality, and alcoholism" (p. 313). These theories have been corroborated in large-scale epidemiological studies.

In a widely recognized study of the structure of psychopathology, Kruger et al. used confirmatory factor analysis to describe the patterns of comorbidity; the best fitting model classified disorders into higher-order factors of externalizing (EXT) and internalizing (INT) disorders (1999). This model is well replicated and has been shown to be valid in a large meta-analysis (n= 23, 557) of five populations-based studies of comorbidity (Kruger & Markon, 2006). In the EXT model of psychopathology, the following classes of disorders (as defined by the Diagnostic Statistical Manual 5; American Psychiatric Association, 2013) are considered to be central components: substance-related and addictive disorders, childhood neurodevelopmental disorders or disruptive, impulse-control and conduct disorders (i.e. attention deficit/ hyperactivity disorder [ADHD], conduct disorder [CD], & oppositional defiant disorder [ODD]), and personality disorders in adulthood (i.e. antisocial personality disorder [ASPD]). Many of these models use symptoms counts, which meta-analyses suggest improves power, reliability, and validity over models of discrete diagnoses (Markon, Chmielewski, & Miller, 2011). While BD factors often use the same psychopathology symptoms, additional measures in line with the definition used by Sher and Trull (1994; i.e. novelty seeking, impulsivity or lack of constraint) have been included. Such

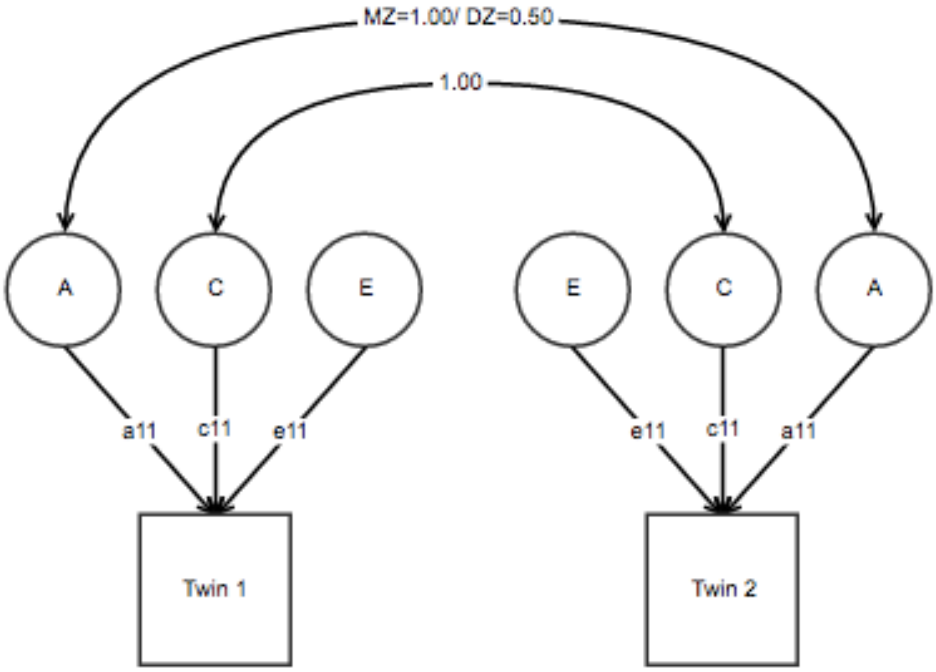
measures include (lack of) constraint (Krueger et al., 2002), novelty seeking (Young et al., 2000; 2009), impulsivity and aggression (Vrieze et al., 2013), and behaviors that are non-normative developmentally (e.g. early sexual behavior; McGue & Iacono, 2005; McGue, Iacono, & Krueger, 2006b; Hicks et al., 2011; Vrieze et al., 2013). Models of EXT and BD have often been used interchangeably. Indeed, BD has also been described as a general liability toward EXT behaviors or disorders (Young et al., 2000; 2009).

Biometrical Modeling using Twin and Family Studies

Much of what is known about the structure of BD comes from biometrical modeling. Twin, adoption, and family models decompose variance of a trait or underlying latent factor such as BD into additive genetic, or sum effect of each individual segregating allele (A), dominance genetic or interactive effects of alleles within loci (D), shared or common environmental (C), and non-shared environmental (E) factors by leveraging differences in similarity among pairs of relatives sharing differing magnitudes of these influential factors. In path models, standardized squared path estimates represent the proportion of variance explained by each factor and are commonly referred to as influences on the trait (i.e., a^2 represents the proportion of variance explained by additive genetics, also referred to as narrow-sense heritability). Figure 1.1 portrays a twin model, though models using other types of genetically related pairs are common.

Figure 1.1

Univariate twin model with A, C, and E factors



The logic of classical twin modeling rests on several statistical assumptions, though the validity of the equal environment assumption (EEA) is most widely debated. The EEA requires that the magnitude of environmental influences shared between co-twins is the same between monozygotic (MZ) twins (which share 100% of their alleles identical by decent) and dizygotic (DZ) twins (which share, on average, 50% of their alleles identical by decent). It is possible that MZ twins experience greater environmental similarity (e.g., more similar dress or time together) than DZ twins, though the EEA is not violated if these similar environments are uncorrelated with the trait of interest. Importantly, the EEA has been shown to valid for many traits (Loehlin & Nichols, 1976; Conley, Rauscher, Dawes, Magnusson, & Siegal, 2013), and LoParo & Waldman demonstrated increased MZ environmental similarity did not moderate the MZ and DZ twin correlations for childhood externalizing traits (2014).

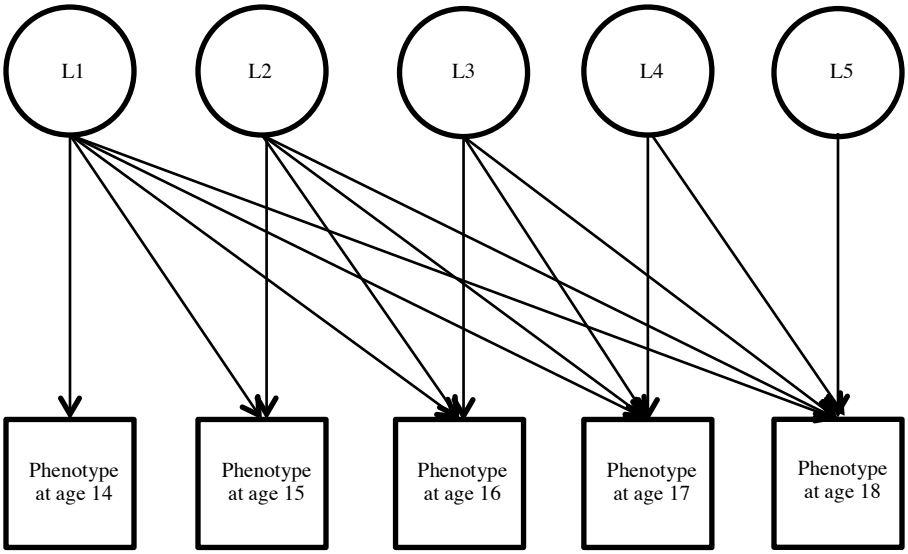
Given that these assumptions are met, an estimate of a^2 and d^2 can be inferred from the extent to which MZ twin pairs are more similar than DZ twin pairs. Finally, c^2 will contribute to similarity in both MZ and DZ pairs, while e^2 (environmental influences uncorrelated across twins or measurement error) will only contribute to trait variance and not twin resemblance. Finally, the similarity of DZ pairs will approach the similarity of MZ pairs as shared environmental influences increase in magnitude. The similarity of DZ pairs will be much less than the similarity of MZ pairs in the presence of strong dominance effects (e.g. MZ twins shared 100% of their dominance effects, while DZ pairs share 25%, on average). With twins reared together there is insufficient information to estimate both d^2 and c^2 in the same model. ACE models (including a^2, c^2, and e^2 estimates) are more appropriate than ADE models (including a^2, d^2, and e^2 estimates) when DZ twin correlations (rDZ) are more than half of the MZ twin correlation (rMZ), suggesting that the presence of shared environment is masking any effects of dominance that may exist.

In adoptive and family studies, estimates are derived using pairs with various genetic relationships. In one simple case, variance can be estimated by comparing the similarity in biological siblings reared in the same home (e.g., who share approximately 50% of additive effects and 25% of dominance effects, on average) to non-genetically related adoptive sibling pairs reared in the same home. Alternatively, variance components can be estimated by comparing similarity between children and their biological parents (e.g., who share approximately 50% of additive effects and 25% of dominance effects, on average) to children and their adoptive parents who are not genetically related. These models can be extended to include family members with various degrees of genetic relatedness (e.g., half- siblings, cousins, aunts, uncles, etc.).

Similar to univariate twin, adoption, and family models, multivariate studies can decompose sources of *covariance* between several traits into A, D, C, and E factors. Several types of multivariate biometric models of genetic and environmental covariance are briefly discussed here including the Cholesky decomposition, independent pathway model, common pathway model, genetic simplex model, and other extended designs (Neale and Cardon, 1992). These are used to model comorbidity across traits or developmental trends of a single trait across repeated measurements (as demonstrated in Chapter II**).** The Cholesky decomposition model is often considered a base model, as it is a full decomposition of the covariance of all measures into A, D, C, and E factors. The first (A, D, C, or E) latent factor (depicted as L1 in Figure 1.2) explains the variance of the first trait and covariance common to all remaining traits, the second explains the remaining covariance between the second and nth factors, etc.

Figure 1.2

Cholesky decomposition model across five repeated measurements

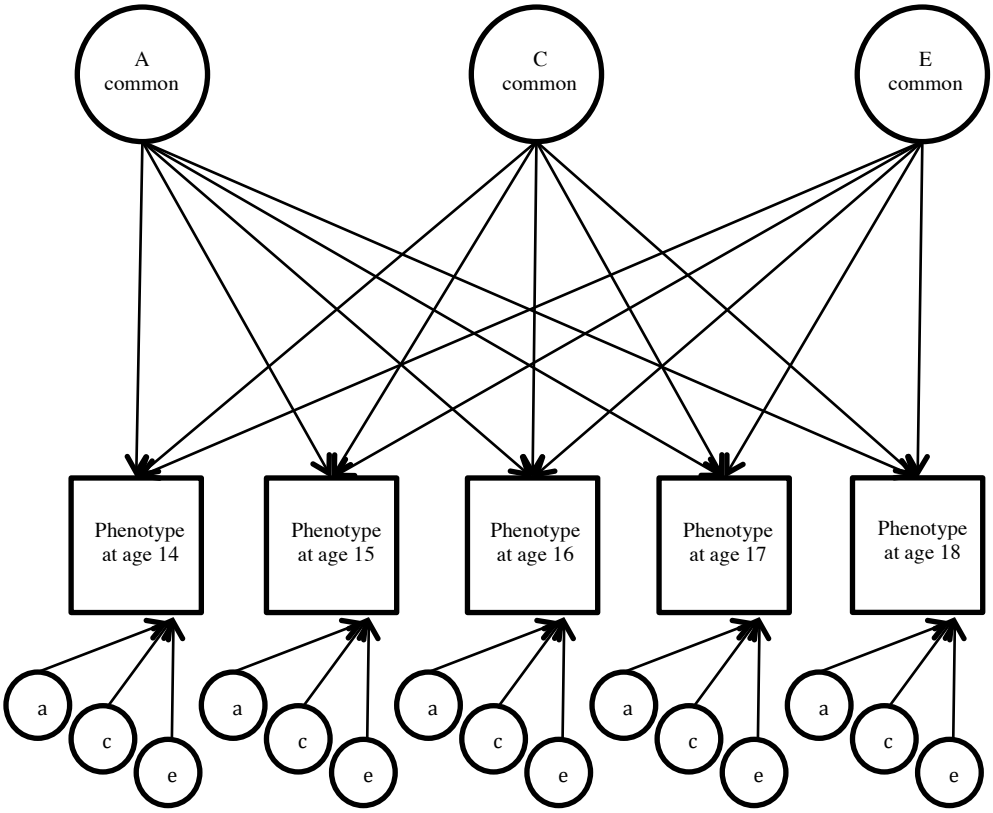


Note: Latent variables (L) can further be decomposed into three separate latent variables reflecting the influences of additive genetics (A), dominance genetics (D), shared environment (C), and non-shared environment (E). Figure 1.2 depicts the model for sibling-1 only; the model for sibling-2 is identical; correlations among the latent variables for the following relationships are fixed as such: MZ (A=1.00, D= 1.00, C=1.00, E=0.00), DZ (A=0.50, D= 0.25, C=1.00, E=0.00), adopted siblings reared together (A=0.00, D= 0.00, C=1.00, E=0.00), control or biological siblings reared together (A=0.50, D= 0.25, C=1.00, E=0.00).

Two more restrictive models include the independent pathway and the common factor model. The independent pathway model estimates general A, D, C, E factors that load on each measurement and allows for measure specific influences (See Figure 1.3). Rather than common A, D, C, E factors that load directly onto each measurement, in the common pathway model these factors load onto a single latent trait that captures covariance across measurement (See Figure 1.4). Similar to the independent pathway model, the common pathway model also allows measure specific influences.

Figure 1.3.

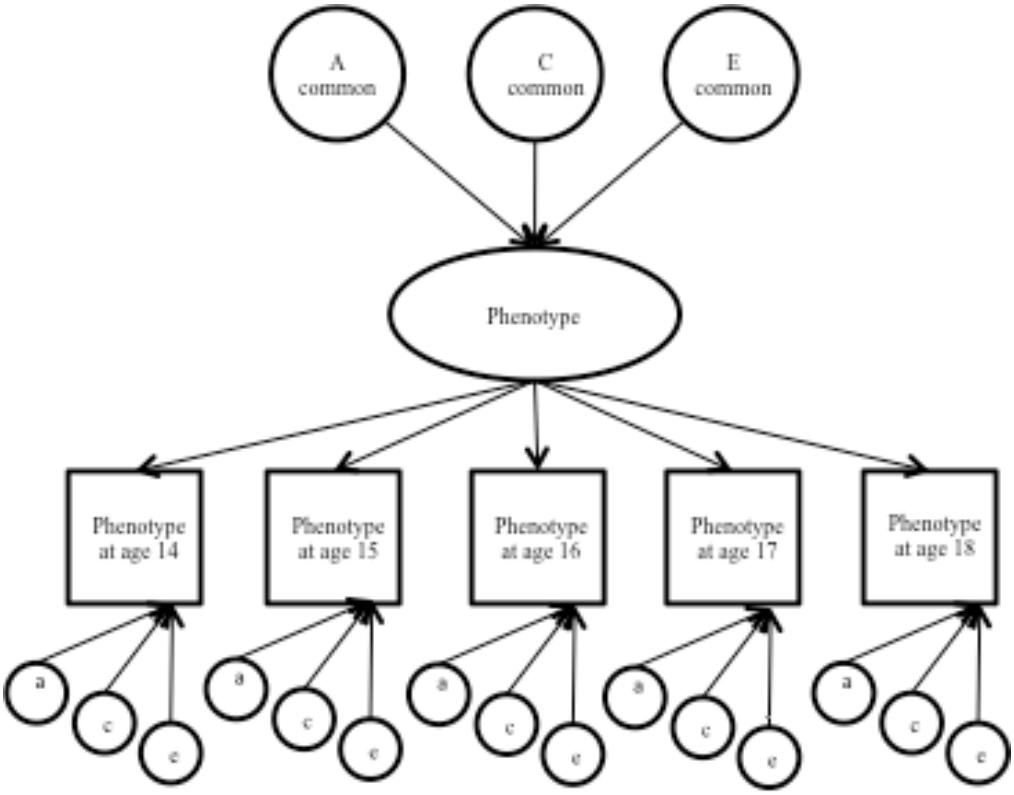
Independent pathway across five repeated measurements



Note: For simplicity of presentation, only the model for sibling-1 is depicted.

Figure 1.4

Common pathway model across five repeated measurements

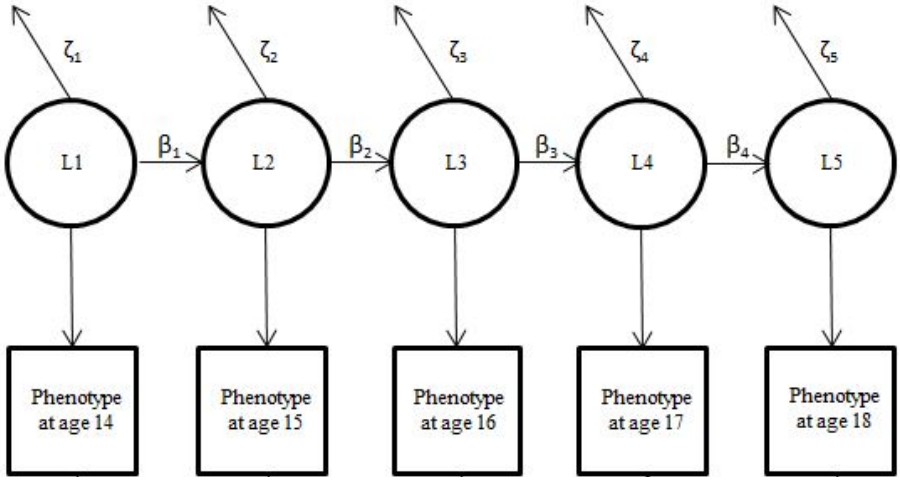


Note: For simplicity of presentation, only the model for sibling-1 is depicted.

The genetic simplex model (Boomsma & Molenaar, 1987) is most appropriate when measuring developmental change (See Figure 1.5), as it parses variance components into those that are transmitted (β; i.e., parameters shared from each time point to the next) and those that are innovative (ζ; i.e., new time point specific influences).

Figure 1.5

Phenotypic simplex model across five repeated measurements



Note: An extension of the phenotypic simplex model is the genetic simplex mode, where latent variables can further be decomposed into three separate latent variables reflecting the influences of additive genetics, shared environment, and non-shared environment. For simplicity of presentation, only the model for sibling-1 is depicted.

Evidence from the biometric models reflect the structure described by epidemiological models, with a single inherited liability best explaining BD. Young et al. found that a highly heritable (a^2= 84%) common factor explained a substantial proportion of the covariance between CD, ADHD, substance experimentation, and novelty seeking in adolescence (2000). A common pathway model best described the covariance of EXT symptoms spanning adolescence and adulthood (i.e. alcohol dependence, adolescent antisocial behavior, CD, measures of constraint, and drug dependence), which yielded a similar heritability estimate (a^2= 81%; Krueger et al., 2002). An underlying factor that explains nicotine dependence, alcohol/drug abuse and dependence, and adult antisocial behavior in early adulthood was also highly heritable ($a^2\approx$ 70%), though a factor explaining childhood problem behaviors including initiating/use of substances, police contact, and sexual intercourse was found to be less heritable ($a^2\approx$ 20% averaging between male and female paths; McGue, Iacono, & Krueger, 2006b). The sum of these

studies suggests that the underlying factor is more heritable than any single component (Iacono, Malone, & McGue, 2008). Additionally, these studies found little or no evidence for significant shared environmental influences on the BD factor. Given that these behaviors cluster within an individual because of genes, common genes may also explain familial resemblance.

Several studies of familial transmission confirmed that these behaviors cluster within families because of a highly heritable underlying vulnerability to EXT disorders. First, Hicks, et al. demonstrated that the familial transmission (e.g. from parents to offspring) of adult antisocial behavior, alcohol dependence, conduct disorder, drug dependence (i.e. adulthood EXT symptoms) was largely explained by a common heritable factor rather than disorder specific transmission (2004). Similarly, a general transmission model showed that the same adulthood EXT factor predicted the clustering of ADHD, CD, and ODD (i.e. childhood EXT symptoms) in the offspring with minimal disorder specific effects (Bornovalova, Hicks, Iacono, McGue, 2010). When modeling variance components for these transmitted factors, heritability was estimated at 80% for the adulthood model of EXT and 81% for the model of EXT including childhood disorders (Hicks et al., 2004; Bornovalova et al., 2010).

In contrast to most previous findings, one extended family design provided evidence for some significant shared environmental influences on BD (c^2= 21%) and slightly lower estimates of additive genetic influences (a^2= 65%; Hicks et al., 2011). However, this study deviated from previous reports in that it 1) included information from both adoptive and biological parents and siblings (which provides direct estimates of shared environment), and 2) used a different conceptualization of BD. Though factor indicators for BD included symptoms of CD and ASPD symptoms, behavior that was anti-social, aggressive or non-normative behavior (e.g. early sexual initiation), it *excluded* substance use and SUDs. The authors refit their model using the same BD

factor described in previous studies (Kruger et al., 2002; Hicks et al., 2008) and found the heritability was 76%, suggesting that the difference in estimates was not likely due to the addition of adoptive pairs but rather the due to difference in measures (Hicks et al., 2011).

Finally, the evidence on whether BD is invariant across males and females is mixed. While some studies were underpowered or did not explicitly test for sex effects (i.e., referred to as sex-limitation, Mathers and Jinks, 1977), it is possible that male and female BD presents in different ways. Hicks et al., (2011) identified significant sex effects on EXT, with males endorsing higher levels of illicit drug use, alcohol consumption and dependence, and higher aggressive and antisocial behavior but found no difference in rates of nicotine use. In addition, while BD behaviors increase from age 17-24, they increase at a greater rate for males (Hicks et al., 2007). Often, these mean differences are controlled for along with other confounds (e.g. age) using regression. Other sources of sex-imitation (i.e. differences in the magnitude or effects of underlying variance components) are typically identified via a model-fitting approach that compares the fit of several nested models (e.g. tests whether constraining male and female parameters significantly reduces the overall fit of the model). In one study investigators could not constrain variance components to be equal across males and females, though the highly similar male and female path estimates suggested that the poor fit was due to variance differences in adolescent and adulthood EXT measures across genders (McGue, Iacono, & Krueger, 2006b). Despite some evidence of mean or variance differences, Krueger (1999) and Kendler, Prescot, Myers, & Neale (2003) found that the *structure* of the EXT-INT model of psychopathology looked the same for males and females. Overall, this provides further evidence for the utility of the BD as a relatively parsimonious model to describe the clustering of impulsive, substance use, and other EXT symptoms within an individual and families.

Given the underlying structure of BD, research (e.g. exploration of causal variants) and intervention efforts benefit greatly from looking at this heritable cluster of behaviors rather than rather than targeting specific behaviors in isolation (Iacono, Malone & McGue, 2008).

Risky Sexual Behaviors and BD

Risky sexual behavior (RSB) increases the risk for contracting human immunodeficiency virus (HIV), other sexually transmitted infections and diseases (STI/ STDs), or unwanted pregnancy. The RSB construct can encompass a number of measures including but not limited to early age of sexual initiation, number of sexual partners, frequency of condom use, sex under the influence of drugs or alcohol, risky sexual acts (e.g. those with higher rates of transmission of disease or infection) or sex with risky partners (e.g. with higher risk of disease or infection). However, the overall "riskiness" of any given behavior has been widely debated and these phenotypes can be difficult to measure (Fenton, Johnson, McManus, & Erens, 2001; Mercer 2010); issues with measuring sexual risk and defining RSB are discussed at length in Chapter IV.

RSB has been considered both a correlate and predictor of BD. These behaviors are predictors of each other, both concurrently (Bailey, Pollock, Martin, & Lynch, 1999; Donohew et al., 2000; Kahn, Kaplowitz, Goodman, & Emans, 2002; Caminis et al., 2007; Charnigo et al., 2013) and prospectively (Guo et al., 2002; Cooper, Wood, Orcutt, & Albino, 2003; Ramrakha et al., 2007; Wymbs et al., 2013; Holoway, Tilman, & Brewster, 2015; O'Hara & Cooper, 2015; Cha, Masho, Mezuk, 2016; Manhart et al., 2016).

Only a limited number of studies have included measures of RSB when creating factor scores of BD or problem behavior. Primarily, *early* age of sexual initiation has been used to construct latent factors of BD (Hicks et al., 2011; Vrieze et al., 2013) and adolescent problem behavior (McGue & Iacono, 2005; McGue, Iacono, & Krueger, 2006b). It is possible that

including RSB in a BD model changes the underlying covariance structure so that the factor differs in magnitude for genetic and environmental influences compared to studies limiting the BD measure to symptoms of antisocial or EXT psychopathology. Indeed, at least one study that included early sexual intercourse found evidence for a small effect of shared environmental influences (Hicks et al., 2011), which is a non-significant variance component in many other studies of BD. Though there is some evidence that traditional BD measures and RSB co-occur due to environmental reasons, most studies suggest that the covariance is largely genetic in nature.

Several biometrical studies have explicitly tested the nature of the overlap between RSB and BD. Verweij et al. (2009) used 4,904 twins from Australia to model the overlap between a moderately heritable (a^2= 34%) composite of RSB (i.e., number of lifetime sexual partners, impaired sex, unwanted pregnancy, unprotected sex without wanting to get pregnant, condom-less sex with someone other than a regular partner, STI/STDs, or sex with someone other than regular partner while in relationship, with partner met that day, or with more than one partner in 24 hours) and moderately heritable (a^2= 56%) composite of adolescent misconduct (e.g., sum score of ever drank alcohol, got drunk, smoked cigarettes, smoked marijuana, used other illegal drugs, stole, vandalized property, cheated, lied, etc.). In the full model, genetic influences explained 61% of the phenotypic correlation (r=.50) and shared environmental influences explained 27%, though the shared environmental overlap was not statistically significant (Verweij et al., 2009). Similarly, Samek et al. (2014) used a Cholesky decomposition to model the covariance between age 14 BD, age of sexual initiation, and early adulthood RSB in 1,512 participants of the Minnesota Twin Family Study. Results showed that the liability to age 14 BD was shared with later sexual behaviors and explained part of the correlation between the two; for

males in particular, this was explained by common genetic influences on BD (Samek et al., 2014). Thus, the current literature suggests that the association between RSB and BD can be explained by both common genetic and environmental influences (the latter source of variance may be more significant in adolescence).

Understanding the shared etiology between BD and RSB has direct consequences for sexual health policy and RSB prevention programs. Several studies have tested causal assumptions about BD and RSB using genetically informed within family designs (i.e. discordant twin design, co-twin control design, sibling or other within family analyses). Monozygotic (MZ) twins provide an opportunity for a natural quasi-experimental design, where co-twins can be used as their twin's control (i.e. they are perfectly matched on genetic effects, and share family-level environmental influences). Dizygotic (DZ) twins and siblings can be thought of as less stringent, but still highly valuable, controls. Causal assumptions can be tested by comparing outcomes for pairs discordant on a predictor or by comparing within-family to between-family effects (Rutter 2007; Vitatro, Brendgen, Arseneault, 2009; McGue, Olser, Christensen, 2010). In addition to controlling for shared genetic and environmental factors between twins, additional potential mediators between the predictor and outcome can be controlled for. Several examples in the literature exist, for both looking at the causal effects of predictors of RSB and to explore potential causal effects of RSB on presumed outcomes.

Of particular relevance, these methods have been applied to test the casual assumptions of various predictors of RSB. These include delinquency (e.g., graffiti and stealing, etc.; Rowe, Rodgers, Meseck-Bushey, & St. John, 1989; Harden et al., 2008), father's absence (Mendle et al, 2009), cognitive impairment and poor academic achievement (Harden & Mendel, 2011a; Garrison & Rodgers, 2016), early adverse events (i.e., physical and sexual trauma, as well as

adolescent cannabis and cigarette use; Donahue, D'Onofrio, Lichtenstein, & Långström, 2013), early smoking, drinking, drunkenness (Deutsch et al., 2014), and early cannabis use (Agrawal et al., 2016). In most cases, controlling for familial influences (i.e., genetic and shared environmental influences) reduced the association between these predictors and RSB, though causality could not be ruled out entirely. Importantly, these examples are behavioral predictors that are thought to be modifiable. Other studies have found genetic and environmental overlap with constructs that may be less interesting from a public policy perspective (e.g. shared etiology with personality factors; Zietsch et al., 2009).

These methods have also been applied to look at the effects of RSB, particularly to examine how sexual behavior in adolescence predicts maladaptive outcomes. Family studies comparing sisters discordant for early pregnancy have been used to tease apart the effects of predictors of early pregnancy from the effects of teenage child bearing. For instance, the economic gap in adulthood between teenage mothers and sisters is significantly smaller than between teenage mothers and the general population (Hoffman, Foster, & Furstenberg, 1993; Hoffman et al., 1988), suggesting that family level variables account for some adaptive outcomes associated with early pregnancy.

Family studies have also been used to explore or tease apart the consequences of early sexual initiation from genetic and environmental influences on timing of sexual initiation. Using twin pairs, common genetic and environmental influences were identified for adolescent sexual behavior and delinquency (Harden & Mendle, 2011b). Additionally, early sexual initiation (i.e. before age 16) was shown to be an unlikely cause of a number of associated maladaptive outcomes in early adulthood including childbearing by age 20 for women, cigarette and cannabis use, alcohol abuse and dependence, depression symptoms or episode, or criminal offending by

age 25 when controlling for genetic and environmental confounds (Donahue, Lichtenstein, Långström, & D'Onofrio, 2013). In fact, when controlling for genetic and familial factors, *late* sexual initiation may be most associated with poor relationship outcomes in adulthood (Harden 2012). Huibregtse et al. demonstrated that early sexual initiation did not necessarily cause additional RSB in early adulthood, but that shared genetic or environmental predictors influenced both behaviors (2011). As previously mentioned, some of the common influences between age of sexual initiation and RSB in early adulthood were also shared with early (i.e. age 14) BD (i.e. the overlap was primarily genetic for males and due to shared environment for females; Samek et al., 2014). These natural experiments provide valuable insight for public policy, and in many cases suggest that early health interventions may be more effective targeting downstream sexual behavior.

It is possible that genetic and shared environmental influences on BD could account for a substantial amount of covariance between early predictors of RSB (e.g., delinquency or substance use; Zimmer-Gembeck & Helfand, 2008) and RSB, as well as between RSB and maladaptive outcomes. As such, Chapter III uses related methods to explore the association between sex under the influence and number of lifetime sexual partners.

Notably, a criticism of these designs is that MZ twins who are discordant on important phenotypes (e.g. a pair where one twin has early sexual initiation) may be substantively different in behavior compared to singletons (Boardman & Fletcher, 2015). Similarly, estimates from the classic twin model will be biased to the extent that major assumptions are violated (e.g. if there is special twin environment, or if the amount of common environment differs across MZ and DZ twins). Extended family designs (Fulker, 1982; Medland & Keller, 2009) or genome wide

models that estimate heritability using non-related individuals can alleviate some of these concerns.

Beyond Twin and Family Studies

While much of what is known about the structure of BD and RSB was originally discovered using twin and family designs, innovations in statistical genetics and genome wide methods are now being applied to these phenotypes. Such methods are useful in that they 1) do not share the same assumptions as twin and family methods, 2) can be conducted on "unrelated" samples, and 3) provide insight into potential causal genes and underlying biological mechanisms of a trait.

Genome wide association studies (GWAS) have been used to identify genetic variants or markers that are associated with behavioral traits. These methods typically measure single base pair differences across the genome called single nucleotide polymorphisms (SNPs). In large samples of unrelated individuals, SNPs that are associated with the trait after correction for multiple testing are considered genome wide significant hits ($p <= 5.0e-8$). These markers may not necessarily be causal, but may be correlated with nearby causal variation (i.e. in linkage disequilibrium [LD] with a causal variant). At least two GWAS have focused on BD specifically. Using a sample of 7,188 participants from 2,300 families, McGue et al. tested five BD traits (nicotine use, alcohol consumption, alcohol dependence, illicit drug use, and non-substance related BD) separately and found a single SNP association that met genomewide significance criteria (2013). Derringer et al. did not identify a genome wide significant hit for a BD factor in a smaller sample of 1,901 adolescents selected for antisocial behavior (2015). However, there are several reasons why GWAS may not produce significant results despite the high heritability of BD estimated from biometric models.

In recent years there has been much discussion of the so called "missing heritability," or the difference in explained heritability from significant GWAS hits (i.e., h^2_{GWAS}) compared to what is estimated from classic twin, adoption, and family studies (Maher, 2008; Manolio et al, 2009). It is now largely recognized that genetic variation for complex traits is likely explained by many variants each with relatively small effects; thus, many original GWAS samples are underpowered to detect genome-wide significant hits (Lee, Wray, Goddard, & Visscher, 2011).

Several solutions have been proposed and implemented to improve GWAS. The optimal strategy is to create large samples that have consistent genotyping and phenotyping across participants (Wray et al., 2012). Alternatively, several consortium and working groups have been created to utilize and integrate pre-existing GWAS samples. Large-scale GWAS meta-analyses have been successful in identifying GWAS hits that were otherwise undetected in small, underpowered samples. For instance, while earlier studies were considered successful if they detected one or several genome wide significant hits, the Schizophrenia Working Group of the Psychiatric Genomics Consortium has now identified over 100 independent markers by combining many samples (Ripke et al., 2014). The number of genome wide significant hits for BD and sexual behavior phenotypes have also dramatically increased with the creation of large scale meta analyses and consortium studies (Tobacco and Genetics Consortium, 2010; Barban et al., 2016; Day et al., 2016; Stringer et al, 2016) and are likely to increase as large open source studies reach sample sizes of 500,000 individuals (i.e., UK Biobank: Allen, Sudlow, & Peakman, 2014). A final improvement to GWAS has been a renewed emphasis on optimizing phenotypes reducing phenotype heterogeneity (Manchia et al., 2013) and measurement error, which can be in conflict with consortia efforts (Bennett et al., 2011).

Though improvements in GWAS will likely raise the average h^2_{GWAS} in time, it is also possible to estimate heritability using all available markers across the genome. Markers typically are limited to single base pair differences (i.e., single nucleotide polymorphism, or SNPs). Genomic-related-matrix restricted maximum likelihood (GREML) estimates trait heritability using measured genetic similarity based on common SNPs to predict phenotypic similarity in unrelated individuals (Yang et al., 2010). For instance, Vrieze et al. explained 10-30% of the trait variance in several substance use measures and a non-substance related BD trait (2013) using a GREML method called genome wide complex trait analysis (GCTA; Yang, Lee, Goddard, & Vissher, 2011). GCTA can also be extended to estimate the genetic overlap across traits in the same sample, by estimating the genetic correlation (rG).

A related approach called linkage disequilibrium score regression (LD-score regression; Bulik-Sullivan, et al., 2015a) can also estimate heritability from SNPs (h^2_{SNP}) while partitioning genetic signals from environmental confounding (e.g., such as to population stratification). An additional strength of the method is that is that genetic correlations (rGs) can be estimated using GWAS summary statistics, which allows for estimation of rGs with phenotypes collected in other samples (Bulik-Sullivan, et al., 2015b).

A final method of exploring shared genetic etiology is to use polygenic scores (PGSs), which use effect sizes obtained from an existing GWAS (discovery sample) to predict phenotypes in independent samples. As meta-analyses and consortiums continue to increase overall in size, the predictive power of PGSs should also improve (Dudbridge, 2013; Rietveld et al., 2013).

Chapter V uses several of these methods to explore the genetic architecture of RSB and the genetic overlap between RSB, BD related diseases and traits (e.g. smoking, psychiatric,

personality), and other fitness phenotypes. This is an important extension to the twin and family literature, as these methods rely on different sets of assumptions than biometric models. Evidence of shared genetic etiology using relatively new genome wide approaches would provide powerful corroborating evidence.

Aims

This dissertation expands upon the study of BD and RSB by exploring the etiology and overlap between these behaviors using a variety of methods (multivariate biometric models and quantitative genome wide methods).

Chapter II

Cross-sectional twin studies have assessed the genetic and environmental etiologies of substance use during adolescence, though comparisons of results across different samples, measures and cohorts are problematic. The major aim of the study was to add to the existing literature in the following two ways 1) corroborate twin findings with adoption findings (which provide a direct estimate of shared environment, and 2) address limitations (i.e. the majority of twin studies are cross sectional) by conducting a longitudinal study with dense and consistent measurements across adolescence. Thus, this study used a less common adoption design to test developmental trends in substance-related behaviors central to the BD factor.

Chapter III

The aim of this study was to test an underlying causal assumption commonly proposed in the RSB literature: that drinking or using drugs during sex will cause more RSB. While these behaviors are highly correlated, it is possible that they co-occur because they share common underlying vulnerabilities (e.g. similar genetic propensities or environmental influences). With a genetically informed twin design, it is possible to decompose sources of covariance between sex

under the influence and RSB (measured by number of lifetime sexual partners) into A, C, and E effects to identify patterns that are consistent or inconsistent with causality. A second aim is to test whether this overlap is limited to drug use during sex (e.g. drug impaired sex) or outside of sexual contexts (substance use more generally).

Chapter IV

This chapter addresses issues relevant to the definition and measurement of RSB, through a review of the literature and an exploration of survey responses to our primary measure of RSB (i.e., M-RBQ) and to create a revised version. As such, this chapter serves to 1) inform and clarify the meaning and interpretation of the variables selected for analysis in Chapter III (i.e. number of lifetime sexual partners and sex under the influence), 2) assess the utility of the M-RBQ and create a revised version, and 3) to select an optimal phenotype to be used in later genetic analyses (i.e., in Chapter V).

Chapter V

The aims of this chapter are to explore the genetic architecture of number of lifetime sexual partners (an index of RSB) and test the genetic overlap with behavioral disinhibition (BD) and related diseases and traits, as well as other fitness phenotypes. We use several genome wide approaches including exploring top hits from a genome wide association study in the UK Biobank and replication samples, estimating heritability explained by common SNPs estimated through LD score regression, and by exploring genetic overlap between number of lifetime partners across samples and with other related traits.

CHAPTER II

A LONGITUDINAL ADOPTION STUDY OF SUBSTANCE USE IN ADOLESCENCE

Background

The transition from adolescence into adulthood is a particularly formative period for a number of behaviors. In the case of substance use, both initial experimentation and continued use are thought to be due to a combination of genetic and environmental influences. Similar to other phenotypes, it is likely that the magnitudes of these influences vary across time and context. While several twin studies have examined the extent to which genes and environment influence substance use at various ages, differences across samples and measures make the results less interpretable than findings from prospective developmental studies.

An essential aspect of understanding influences on the frequency of substance use behavior is to first look at what motivates trying substances for the first time. Ever having tried a particular substance will herein be referred to as 'use' if tried, and 'no use' if never tried. Estimates of the proportion of genetic and environmental influences on use/no use appear to vary by age of sample. For example, in a sample of male and female twins in adulthood (mid-thirties) the heritability for liability to use tobacco was .73 (Maes et al*.,* 2004). In a younger sample (age 17-18; Han, McGue, & Iacono, 1999), which may not be fully past the "age of risk" (Lopez-Leon & Raley, 2012), the heritability of tobacco use was estimated at .11 (females) and .59 (males), with shared environment estimates of .71 and .18 for females and males, respectively. Parameter estimates for alcohol use were similar in that sample (Han, McGue, & Iacono, 1999). When splitting a twin sample into three age groups (i.e. twins aged 13-15, 16-17, and 18-20), heritability estimates for 'ever' using marijuana declined with age while shared environmental influences increased (Distel et al., 2011). A similar increase in the magnitude of shared

environmental influences was found when comparing 12-14 year old twins to 15-16 year old pairs for initiation of alcohol use, especially among females (Koopsmans, Lorenz, & Boomsma, 1997). Evidence for age-moderated influences suggests that these parameter estimates should be interpreted within the context of specific life stages, in which differential environmental or genetic influences may be of importance. The authors of a recent meta-analysis of twin-studies of marijuana use acknowledged the possible moderating effect of age on estimates of genetic and environmental influences across time, although the findings are limited by the relatively small number of genetically-informative longitudinal samples currently available (Verweij et al., 2010).

Similar developmental issues exist in the literature on the frequency of substance use, where most reported results are also cross-sectional. In a sample of twin pairs ranging from 8 to 16, Maes and colleagues found moderate to high heritabilites for past month substance use (.60, .56, and .27) and a small to moderate proportional influence of shared environment (.18, .17, and .35) for tobacco, alcohol, and marijuana, respectively (Maes et al., 1999). In a combined twin, sibling, and adoptive sample of adolescents (mean age, 15.85, SD, 2.08 years), moderate to high heritabilities for regular tobacco and marijuana use were reported, with no genetic influences on regular alcohol use (Rhee et al., 2003).

While cross sectional studies have been informative, more powerful longitudinal designs measure substance use at several ages or developmental stages, and eliminate the problems associated with cross sample comparisons. In one quasi-longitudinal cross-sectional study, a life history calendar approach was used to bolster retrospective recall of average monthly use for nicotine, alcohol, and marijuana at various life stages (Kendler et al., 2008). Shared environmental influences on frequency of alcohol and marijuana use were important through

adolescence, and genetic influences increased in relative importance into adulthood. For frequency of cigarette use, shared environment influences were only evident for very early use and then declined steadily from age 15 as genetic influences became increasingly important (Kendler et al. 2008). A one-year longitudinal study of the FinnTwin16 cohort found substantial shared environmental influences on alcohol use at age 16 (.79) and 17 (.76), with smaller estimates for frequency of alcohol use across the same time span, (.35 and .22 at ages 16 and 17, respectively; Viken, Kaprio, Koskenvuo, & Rose, 1999). Following the FinnTwin12 and FinnTwin16 cohorts up to age 25, the relative importance of shared environment for females increased while the heritability for the frequency of alcohol use decreased. Estimates for males remained stable from ages 17 to 25 (Pagan et al., 2006). Finally, a longitudinal study tracking smoking, alcohol and illicit drug use across adolescence showed some increase in heritability across ages (Baker et al., 2011).

Like twin studies, adoption designs also capitalize on the varying degrees of genetic similarity of sibling pairs to estimate the extent of genetic and environmental influences on a given trait. Biological sibling pairs reared in the same home, who share on average, 50% of their alleles identical by descent, may be similar on a given phenotype because of shared environment or shared genes. In the absence of selective placement, any similarity between adopted sibling pairs, who are not genetically related, must be attributed to shared environment. Thus, adoption studies can provide a direct estimate of the influence of shared environment on a phenotype—an estimate that can be used as a powerful anchor for comparison with findings from twin studies. Similarly, parent-offspring designs are useful for estimating the magnitude of shared environmental influence by comparing similarity of children to their biological and adoptive parents. Parent-offspring and sibling-based adoption designs differ in several ways, most notably

are the specific sources and magnitude of shared environment. While neither parent-offspring nor adoptive-sibling designs rely on the equal environments assumption of twin studies, there are also notable differences in the source and magnitude of shared environmental effects between twins and siblings. For example, because twins are the same age they tend to spend more time together than non-twin siblings. Sibling-based adoptive designs can be influenced by factors such as test age differences between adoptive and biological sibling groups. While these design differences could lead to slightly different estimates, the comparison is still warranted.

While there have been a few cross-sectional adoption studies that have investigated substance use at specific points during adolescence (McGue, Sharma, & Benson, 1996; Buchanan, McGue, Keyes, & Iacono, 2009), and one recent parent-offspring longitudinal study (McGue, Malone, Keyes, & Iacono, 2014), no sibling based adoption study has investigated the stability or change of these influences from adolescence into adulthood.

The current study had several aims. We sought to corroborate previously described estimates of biometrical parameters based on twin research using an adoptive sample, which provides a direct estimate of shared environmental influences common to siblings. Further, as the first comprehensive longitudinal sibling-based adoption study of substance use spanning adolescence to early adulthood (i.e., age 18), we examined whether the estimates of heritability and environmental influences change as adolescents transition through significant biological or socio-environmental life stages. Finally, we tested a series of biometrical models to determine the extent to which the change in estimates over time is due to stable or novel genetic and environmental influences. We were particularly interested in the transition from adolescence to early adulthood (i.e., age 18) as changing cultural attitudes, increased independence, and changes

in legal rights (e.g., ability to legally purchase cigarettes) may underlie important environmental changes during this time.

Methods

Sample

Participants were from the Colorado Adoption Project (CAP), a longitudinal study following adoptive children, matched controls, and their families (Plomin & DeFries, 1983, 1985) approximately yearly from infancy into adulthood. Adoptive probands were ascertained through two Denver adoption agencies, while control probands were recruited from hospitals and matched to adoptive families based on sex of proband, number of children in the family, age and occupation of father, and father's years of education. Enrollment in the CAP occurred between 1976 and 1983, and resulted in a final sample of 245 adoptive families and 245 matched control families (Rhea, Bricker, Wadsworth, & Plomin, 2013). The most proximal younger sibling of the proband (if available) was also recruited into the study as they reached the age of the proband at first assessment, so that sibling pair similarity could be compared across adoptive and control families. While proband assessments at any age (e.g., age 14) generally clustered within a given year, there was variation in the birth years of the siblings tested at a given age. Siblings were also assessed approximately annually, so that it was possible to compare measures taken when both the proband and the sibling were at a given age (e.g., 14). In contrast, cross-sectional studies compare sibling similarity within a given test year (e.g., proband at age 14, sibling at age 11) when influences on substance use may vary in both source and magnitude. (For further details of the CAP recruitment and assessment protocols, refer to Rhea et al., 2013). Table 2.1 shows the number of control and adoptive sibling pairs tested at each age.

Table 2.1.

Descriptive statistics at each age

	14	15	16	17	18
Control					
# of Pairs	92	69	97	87	58
Proband Age	14.52±.38	15.38±.33	16.34±.56	17.46±.40	18.37±.26
Sibling Age	14.44±.33	15.37±.31	16.40±.58	17.49±.38	18.26±.50
Age Diff	.37±.26	.32±.27	.43±.58	.42±.38	.39±.25
Adoptive					
# of Pairs	77	54	76	77	40
Proband Age	14.51±.41	15.39±.26	16.29±.41	17.52±.35	18.48±.28
Sibling Age	14.52±.37	15.38± .29	16.44±.58	17.51±.35	18.23±.53
Age Diff	.44±.30	.28±.23	.41±.53	.44±.47	.52±.25

The CAP includes early and frequent interviews for substance use, biannually from ages 12-18. Due to low prevalence of any substance use in early adolescence, we began analysis with the age 14 assessment. We used data from probands and siblings who were tested at the same age (i.e., age at time of assessment of the sibling was within one year of the proband's test age; see Table 2.1). When individuals had multiple assessments within a year (starting at age 15), we selected those assessments that would minimize the test age gap within sibling pairs. Although we used identical procedures for adoptive and control families, there was a trend (in 3 out of 5 waves) for the mean difference between the test age of a proband and his/her sibling to be greater in adoptive families compared to control families. These mean differences were small and generally not significant, with the exception of the age 18 assessment (adoptive age difference [M=.39, SD=.25], control age difference [M=.52, SD=.25], $t(96)$=2.53, p=.013]). Though significant, this difference corresponds to a mean test age difference of approximately 50 days at the age 18 assessment.

Measures

Substance use was assessed with both *use/no use* and a measure of *quantity/ frequency of use.* We assessed cigarette use, alcohol use, and marijuana use since they were the most commonly endorsed drugs used from age 14-18.

Use/no use was coded as a dichotomous variable ("no/never" [0], "yes" [1]) based on the questions: "Have you ever smoked cigarettes?", "Have you ever had a drink of beer, wine, or liquor?", and "Have you ever tried marijuana?". Prevalence of use of cigarettes, alcohol and marijuana at each age are shown in Table 2.2.

Table 2.2.

Prevalence of having tried substances across age

	14	15	16	17	18
Control					
Cigarettes					
Proband	36.8	51.6	55.9	57.5	69.0
Sibling	50.0	57.7	62.9	70.1	79.3
Alcohol					
Proband	31.1	54.3	61.8	90.8	87.9
Sibling	46.7	62.8	68.0	92.0	89.7
Marijuana					
Proband	16.0	24.5	33.3	43.7	53.4
Sibling	23.8	28.2	38.1	56.3	60.3
Adoptive					
Cigarettes					
Proband	42.9	59.7	67.7	63.3	80.0
Sibling	40.7	51.4	54.5	68.8	65.0
Alcohol					
Proband	48.1	62.9	64.2	80.5	87.5
Sibling	45.9	56.9	56.8	92.9	82.5
Marijuana					
Proband	15.4	24.5	35.3	59.7	55.0
Sibling	18.6	31.0	35.2	54.5	55.0

Quantity/ frequency of use items varied across substance and age of test. Cigarette *quantity/ frequency of use* was assessed with the question "How frequently have you smoked cigarettes during the past 30 days?," and was coded on a seven point scale ("None" [0], "Less than 1 cigarette a day" [1], "1-5 cigarettes a day" [2], "½ pack a day" [3], "1 pack a day" [4], "1 ½ packs a day" [5], or "2 packs a day" [6]). Alcohol and marijuana *quantity/ frequency of use* were assessed on 6 or 12 month scales and were converted to a month long seven point scale for consistency ("0 times" [0], "1-2 times" [1], "3-5 times" [2], "6-9 times" [3], "10-19 times" [4], "20-39 times" [5], or "40 or more times" [6]). Raw scores were corrected for sex and log transformed to minimize skewness. Models were conducted using standardized scores. Notably, Table 2.3 shows a general trend of increasing means and standard deviations for quantity/frequency across ages. This is consistent with the increasing prevalence of use across age in Table 2.2.

Table 2.3.

Mean and standard deviation quantity/frequency of use at each age (raw scores)

	14	15	16	17	18
Control					
Cigarettes (n)	92	68	94	87	55
Proband	1.15±.57	1.15±.63	1.28±.81	1.68±1.18	1.80±1.45
Sibling	1.12±.39	1.29±.96	1.35±.92	1.63±1.05	1.75±1.42
Alcohol (n)	92	65	95	86	57
Proband	1.11±.46	1.29±.68	1.52±1.06	1.99±1.36	2.07±1.05
Sibling	1.17±.57	1.49±.90	1.58±.91	2.15±1.38	2.40±1.22
Marijuana (n)	92	68	95	87	57
Proband	1.08±.37	1.15±.60	1.27±.96	1.55±1.34	1.39±.84
Sibling	1.14±.66	1.29±.99	1.32±1.02	1.84±1.68	1.63±1.36
Adoptive					
Cigarettes (n)	75	53	76	76	40
Proband	1.24±.75	1.36±.83	1.50±1.08	1.92±1.41	2.10±1.44
Sibling	1.40±1.03	1.62±1.18	1.58±1.92	2.22±1.55	2.10±1.48
Alcohol (n)	74	53	76	74	40
Proband	1.28±.80	1.40±.79	1.64±1.13	2.19±1.50	2.30±1.22
Sibling	1.16±.52	1.40±.72	1.42±.84	2.08±1.18	2.10±1.10
Marijuana (n)	75	52	75	76	39
Proband	1.09±.52	1.21±.98	1.38±1.05	1.57±1.46	1.62±1.46
Sibling	1.07±.41	1.11±.38	1.22±.70	1.81±1.66	1.51±1.35

Note: quantity/frequency of use was measured on a 7-point scale (1=0 times, 2=1-2 times, 3=3-5 times, 4=6-9 times, 5=10-19 times, 6=20-39 times, 7=40 or more times) in the past month. Table entries include the Ns, and mean ± standard deviation.

For the analysis of dichotomous use/no use data, potential prevalence differences in substance use conditional on age, sex and adoptive status were accommodated by estimating thresholds separately for adoptive versus control sibling pairs, and at each age. As seen in Table 2, there are strong age trends in the prevalence of use with greater use at older ages. There is also a trend (though less strong) for a higher prevalence of use among adoptive probands compared to nonadoptive probands. No significant sex differences in prevalence were observed across this age range.

Quantity/frequency data were transformed to minimize skewness. Within each subgroup (e.g. control probands, adoptive probands, control siblings, and adoptive siblings), we regressed quantity/frequency scores on sex and obtained residuals. A constant of 5 was added to each standardized residual to remove negative values, and the residuals were then log transformed to minimize the positive skew. Finally, log-transformed scores were standardized to facilitate interpretation of model parameter estimates. For descriptive purposes, raw scores are reported in Table 3. However, all biometrical analyses were conducted on standardized, transformed scores.

Analyses

Descriptive statistics and Pearson's product moment correlations quantifying sibling resemblance for quantity/frequency of substance use in the past month were calculated using the Statistical Package for Social Sciences (SPSS Inc., 2010). Genetic analyses were conducted using the software package Mx (Neale, 1997). Tetrachoric (sibling pair) correlations for substance use/no use were computed allowing for separate thresholds for probands and siblings, adoption status, and different thresholds for each assessment age.

Biometrical models accounted for the genetic covariance structure implicit in the adoption design. Briefly, the covariance between control/biological siblings at a given time point can be parsed into additive genetic influences (a^2), and common environmental influence (c^2). Within adoptive sibling pairs, phenotypic similarity can only be due to common environmental influence in the absence of selective placement. Non-shared environmental influences (e^2) only contribute to the overall variance in a trait in a population; the total variation in the population is assumed to be the sum of a^2, c^2 and e^2. Due to sparse data issues, it was not possible to fit multivariate models to the longitudinal 'use' data. Although we fit models to raw data, many of the 10x10 tetrachoric matrices (proband five waves x sibling five waves for each substance) were

not positive definite. For both adoptive and sibling pairs, some cells of the matrices were empty or yielded correlations of ± 1.0. For this reason only univariate models for use/no use were conduced for the three substances at each of the 5 time points.

For multivariate models, a series of nested models were compared for goodness of fit using standard chi-square difference tests (e.g., Neale & Cardon, 1992). A basic Cholesky decomposition was used as a base model (See Figure 1.1; Neale & Cardon, 1992). Since these models are a full decomposition of the variance-covariance matrix across all measurement occasions, they will necessarily provide a good fit to the data structure (i.e., the Cholesky decomposition is just-identified). Subsequent models were considered to have good fit if the additional parameter constraints did not result in a significant decrement in fit compared to the model fit of the corresponding Cholesky decomposition.

The independent pathway model estimates additive genetic (A), shared environmental (C), and non-shared environmental (E) factors that are common across all time points, as well as age-specific influences (or residuals) that only explain variation at specific measurement occasions (See Figure 1.2) These models allow the common genetic and environmental factors to influence the measured traits to different extents. Age-specific influences also may reflect important developmental changes across adolescence, such as novel influences coming "on-line" at older ages.

Several constraints were added to the general independent pathway model to empirically test developmental trends. Specifically, we tested whether 1) all age-specific influences were significant and 2) whether the common influences affected each age to the same degree or whether the magnitude of these influences increase/decrease across adolescence.

Results

Sibling Correlations for "use/no use"

Table 2.4 shows estimated tetrachoric sibling pair correlations for control pairs (who share both genetic and environmental influences) and adoptive pairs (who share only environmental influences) at assessment ages 14 through 18. Across these ages there was a consistent trend where control sibling pairs were more highly correlated for substance use than adoptive sibling pairs (see Table 2.4).

Table 2.4.

Tetrachoric sibling correlations and univariate parameter estimates for use/no use at each age

	14	15	16	17	18
Cigarettes					
$r_{control}$	.58	.22	.30	.32	.13
$r_{adoptive}$	.27	.09	.24	.31	.05
a^2	.32 (.00-1.00)	.26 (.00-1.00)	.11 (.00-1.00)	.03 (.00-.97)	.16 (.00-1.00)
c^2	.27 (.00-.56)	.09 (.00-.42)	.25 (.00-.49)	.31 (.00-.53)	.05 (.00-.44)
e^2	.41 (.00-.87)	.65 (.00-1.00)	.64 (.00-.95)	.66 (.00-.93)	.79 (.00-1.00)
Alcohol					
$r_{control}$	.40	.36	.64	.45	.75
$r_{adoptive}$	-.08	.12	.17	.16	.43
a^2	.80 (.00-1.00)	.60 (.00-1.00)	.76 (.01-1.00)	.58 (.00-1.00)	.54 (.00-1.00)
c^2	.00 (.00-.28)	.06 (.00-.45)	.24 (.00-.54)	.16 (.00-.53)	.45 (.00-.84)
e^2	.20 (.00-.89)	.34 (.00-1.00)	.00 (.00-.56)	.26 (.00-1.00)	.00 (.00-.74)
Marijuana					
$r_{control}$	.61	.53	.41	.54	.54
$r_{adoptive}$	.15	.19	.22	.15	-.02
a^2	.32 (.00-1.00)	.67 (.00-1.00)	.46 (.00-1.00)	.77 (.00-1.00)	.98 (.00-1.00)
c^2	.17 (.00-.52)	.19 (.00-.59)	.18 (.00-.50)	.15 (.00-.49)	.02 (.00-.45)
e^2	.51 (.00-1.00)	.14 (.00-.88)	.36 (.00-.89)	.08 (.00-.75)	.00 (.00-.79)

Univariate Estimates for "use/no use"

Although confidence intervals are quite broad due to the dichotomous nature of the data and the limited samples sizes at each age, the point estimates suggest substantial genetic influences (a^2) on the liability to use alcohol and marijuana, but only modest effects on cigarette use/no use. Shared environmental (c^2) estimates suggest small to moderate influence of the shared environment across substances and across ages. However, for alcohol use, there is some evidence for increasing shared environmental influences from age 14 to age 18.

Sibling Correlations for "quantity/frequency"

Again, with a few exceptions (e.g., the youngest ages) control sibling pairs were generally more highly correlated for quantity/frequency of substance use than adoptive sibling

pairs. Cross time point correlations were generally more strongly correlated with proximal time points compared to more distal ones. Table 2.5 shows the full proband-sibling correlation matrix across the five time points. Adoptive proband-sibling correlations are shown above the diagonal and control proband-sibling correlations below.

Table 2.5.

Correlations for quantity/frequency of use in past month at each age

	Proband					*Sibling*				
	14	15	16	17	18	*14*	*15*	*16*	*17*	*18*
Cigarettes										
Proband14	1.00	**. 72**	**.60**	**.50**	**.06**	**.07**	**.04**	**.18**	**-.16**	**.06**
15	.92	1.00	**.42**	**.28**	**.03**	**.08**	**.15**	**-.07**	**-.06**	**-.17**
16	.53	.66	1.00	**.38**	**.42**	**.21**	**.13**	**.17**	**-.18**	**.08**
17	.27	.25	.67	1.00	**.67**	**.21**	**.26**	**.16**	**.06**	**.19**
18	.06	-.07	.52	.66	1.00	**-.01**	**.19**	**.23**	**-.29**	**-.01**
Sibling 14	-.06	-.06	.20	.50	.24	1.00	**.80**	**.56**	**.51**	**.40**
15	-.05	-.06	.45	.27	.27	.84	1.00	**.71**	**.60**	**.30**
16	-.05	-.06	.40	.33	.37	.45	.63	1.00	**.52**	**.52**
17	.18	-.01	.43	.41	.45	.39	.40	.60	1.00	**.66**
18	.11.	-.09	.28	.46	.25	.48	.63	.49	.78	1.00
Alcohol										
Proband14	1.00	**.57**	**.53**	**.31**	**.27**	**.26**	**-.02**	**.12**	**.17**	**.22**
15	.43	1.00	**.52**	**.41**	**.43**	**.02**	**.13**	**.04**	**.23**	**-.27**
16	.41	.68	1.00	**.41**	**.45**	**.14**	**.11**	**-.01**	**.20**	**-.18**
17	.36	.18	.55	1.00	**.50**	**-.18**	**.12**	**.04**	**.18**	**-.22**
18	.29	.30	.43	.71	1.00	**-.10**	**-.10**	**-.06**	**.19**	**-.10**
Sibling 14	.02	.16	.22	.52	.27	1.00	**.42**	**.50**	**.26**	**.32**
15	.25	.34	.41	.39	.50	.60	1.00	**.66**	**.71**	**.50**
16	.12	.22	.41	.19	.39	.42	.57	1.00	**.56**	. **42**
17	.16	.30	.27	.26	.28	.27	.46	.34	1.00	**.55**
18	-.12	.00	.24	.24	.39	.08	.37	.49	.39	1.00
Marijuana										
Proband14	1.00	**.10**	**.31**	**.19**	**-.09**	**-.09**	**-.13**	**-.09**	**-.01**	**-.10**
15	.69	1.00	**.55**	**.51**	**.62**	**-.06**	**.20**	**.15**	**-.09**	**.26**
16	.62	.78	1.00	**.47**	**.75**	**-.12**	**.09**	**-.01**	**-.14**	**.07**
17	.35	.35	.25	1.00	**.65**	**.27**	**-.06**	**-.04**	**-.06**	**.01**
18	.34	.43	.40	.79	1.00	**-.01**	**.54**	**.24**	**-.15**	**.13**
Sibling 14	.08	-.07	.42	.10	.20	1.00	**-.07**	**.40**	**-.07**	**-.10**
15	.54	.32	.27	.21	.33	.18	1.00	**.19**	**.19**	**.52**
16	.33	.32	.19	.33	.40	.24	.69	1.00	**.28**	**.23**
17	.38	.36	.17	.27	.41	.19	.45	.45	1.00	**.76**
18	-.08	-.09	-.03	.11	.20	.34	.50	.32	.59	1.00

Note: **Bold** indicates adoptive family correlations, normal typeface indicates control family correlations. Within-proband correlations are in top left quadrant, sibling-proband correlations are in bottom left quadrant (control) and top right quadrant (adoptive), and within- sibling correlations are in bottom right quadrant.

Multivariate Biometrical Results

We reported raw scores for substance use quantity/frequency in Table 2.3 to illuminate several trends (e.g. increasing means and variances across ages). However, substance use variables were log-transformed and standardized prior to multivariate biometrical analysis so that path loadings across ages could be interpreted on the same scale. Unfortunately, sparse data issues, though not as severe as with our use/no use data, precluded fitting simplex models to the longitudinal data. It was necessary to utilize Cholesky Decomposition and Independent Pathway Models which are more robust to sparse data issues.

Model fitting comparisons are presented in Table 2.6. Compared to the base Cholesky decomposition (Model 1), the more parsimonious independent pathway model (Model 2) did not result in a significant decrement of fit for quantity/frequency of use of cigarettes, alcohol, or marijuana—assessed at five measurement occasions. Thus, we used the independent pathway as the base model for subsequent model comparisons to explore possible developmental trends.

Table 2.6.

Model comparisons of biometrical models for five ages with standardized variables

	Model	-2LL	df	AIC	BIC[a]	Model Comparison	Δ -2LL	Δdf	p-value
Cigarettes	1) Cholesky Decomposition	3462.05	1397	668.05	215.94	--	--	--	--
	2) Independent Pathway	3475.05	1412	651.05	206.17	2 vs. 1	13.00	15	.60
	3) IP- Drop A Specifics	3478.54	1417	644.54	202.49	3 vs. 2	3.49	3	.32
	4) IP- Drop C Specifics	3475.31	1417	641.31	200.88	4 vs. 2	0.26	3	.97
	5) IP- Drop E Specifics	3505.32	1417	671.32	215.87	5 vs. 2	30.27	3	<.01*
	6) IP- Equate A common	3484.18	1416	652.18	206.40	6 vs. 2	9.09	4	.06
	7) IP-Equate C common	3477.48	1416	645.78	203.05	7 vs. 2	2.43	4	.66
	8) IP-Equate E common	3506.02	1416	674.02	217.75	8 vs. 2	30.97	4	<.01*
Alcohol	1) Cholesky Decomposition	3755.07	1393	969.07	366.79	--	--	--	--
	2) Independent Pathway	3759.65	1408	943.65	352.81	2 vs. 1	4.58	15	.99
	3) IP- Drop A Specifics	3759.75	1413	933.75	347.44	3 vs. 2	0.10	3	.99
	4) IP- Drop C Specifics	3759.83	1413	933.83	347.48	4 vs. 2	0.18	3	.98
	5) IP- Drop E Specifics	3827.97	1413	1001.97	381.55	5 vs. 2	68.32	3	<.01*
	6) IP- Equate A common	3767.75	1412	943.75	352.52	6 vs. 2	8.10	4	.09
	7) IP-Equate C common	3764.16	1412	940.16	350.73	7 vs. 2	5.51	4	.24
	8) IP-Equate E common	3771.37	1412	947.37	354.33	8 vs. 2	11.72	4	.02*
Marijuana	1) Cholesky Decomposition	3963.09	1463	1037.09	394.88	--	--	--	--
	2) Independent Pathway	3976.57	1478	1020.56	387.98	2 vs. 1	13.48	15	.57
	3) IP- Drop A Specifics	3976.58	1483	1010.58	385.28	3 vs. 2	0.00	3	>.99
	4) IP- Drop C Specifics	3976.61	1483	1010.61	379.95	4 vs. 2	0.04	3	.99
	5) IP- Drop E Specifics	4034.02	1483	1068.02	408.66	5 vs. 2	57.45	3	<.01*
	6) IP- Equate A common	3985.10	1482	1021.10	379.93	6 vs. 2	8.53	4	.07
	7) IP-Equate C common	3979.62	1482	1015.62	382.54	7 vs. 2	3.05	4	.55
	8) IP-Equate E common	3993.32	1482	1029.32	289.39	8 vs. 2	16.75	4	<.01*

Note: [a] sample size adjusted BIC.

As a test of the significance of age-specific sources of variance, we compared a series of models where either the additive genetic (A), shared environmental (C), or non-shared environmental (E) *specifics* were dropped from the base independent models (Models 3-5).

Specifics were dropped independently (e.g. Model 3 dropped additive genetic specifics while shared environmental and non-shared environmental specifics remained in the model). Across all substances, there was a significant decrement in fit only when dropping the age-specific non-shared environmental variance components (Model 5). There were no significant age-specific additive genetic or shared environmental influences. Although we had limited power, it can be seen from Table 7 that the point estimates for specific A and C, with few exceptions, are small and quite often zero.

To test the stability of common influences, we also tested a series of models where the common additive genetic, shared environmental, or non-shared environmental pathways were constrained to be equal (Models 6-8). Across all substances, the additive genetic and shared environmental influences could be constrained to be equal; indicating substantial stability across adolescence. However, some caution in interpretation is warranted given power issues. Non-shared environmental pathways across ages were the most variable and could not be constrained to be equal across age for all three substances.

Standardized parameter estimates and confidence intervals for the base (full ACE) independent pathway models for each substance are shown in Table 2.7. The total proportion of variance explained by additive genetic (a^2), shared environmental (c^2), and non-shared environmental (e^2) factors (i.e. common plus specific influences combined) are also reported.

Table 2.7.
Standardized variance estimates, standardized path coefficients, 95% confidence intervals for independent pathway results (Model 2)

	14	15	16	17	18
Cigarette					
A_{common}	.45(.13,.61)	.44(.06,.64)	.57(.34,.74)	.84(.66,1.00)	.81(.64,.97)
C_{common}	.26(.00,.46)	.37(.00,.58)	.26(.00,.53)	.34(.00,.55)	.16(.00,.45)
E_{common}	.73(.62,.87)	.90(.80,1.00)	.42(.29,.60)	.03(.00,.32)	.01(.00,.29)
$A_{specific}$	.00(.00,.47)	.00(.00,.23)	.55(.00,.74)	.00(.00,.43)	.00(.00,.43)
$C_{specific}$	.00(.00,.19)	.00(.00,.22)	.16(.00,.36)	.00(.00,.24)	.00(.00,.25)
$E_{specific}$	.53(.26,.58)	.00(.00,.27)	.35(.00,.63)	.48(.23,.62)	.59(.40,.72)
a^2	.19	.17	.62	.67	.64
c^2	.06	.12	.09	.11	.02
e^2	.75	.71	.29	.22	.34
Alcohol					
A_{common}	.52(.03,.87)	.46(.02,.63)	.56(.20,.71)	.66(.40,.84)	.77(.57,1.00)
C_{common}	.31(.05,.51)	.46(.17,.64)	.23(.00,.44)	.25(.00,.44)	.00(.00,.34)
E_{common}	.45(.25,.68)	.54(.33,.81)	.46(.25,.72)	.00(.00,.35)	.00(.00,.35)
$A_{specific}$	.00(.00,.38)	.00(.00,.53)	.05(.00,.50)	.00(.00,.57)	.28(.00,.70)
$C_{specific}$	.17(.00,.39)	.00(.00,.33)	.00(.00,.26)	.00(.00,.35)	.00(.00,.37)
$E_{specific}$	.74(.60,.84)	.56(.00,.69)	.65(.39,.75)	.70(.46,.82)	.55(.00,.76)
a^2	.24	.21	.32	.44	.69
c^2	.11	.29	.05	.06	.00
e^2	.65	.51	.63	.50	.31
Marijuana					
A_{common}	.48(.32,.63)	.60(.35,.80)	.73(.55,.90)	.35(.08,.60)	.15(.00,.58)
C_{common}	.00(.00,.31)	.38(.00,.60)	.16(.00,.40)	.22(.00,.47)	.34(.00,.57)
E_{common}	.06(.00,.27)	.24(.00,.43)	.19(.00,.43)	.60(.41,1.00)	.88(.41,1.00)
$A_{specific}$	.00(.00,.48)	.00(.00,.66)	.00(.00,.31)	.00(.00,.54)	.00(.00,.51)
$C_{specific}$	.00(.00,.28)	.00(.00,.35)	.00(.00,.21)	.08(.00,.33)	.12(.00,.40)
$E_{specific}$	.87(.73,.96)	.69(.31,.79)	.62(.49,.73)	.69(.00,.77)	.00(.00,.70)
a^2	.23	.35	.54	.12	.02
c^2	.00	.14	.03	.05	.14
e^2	.77	.51	.43	.83	.84

Note: a^2, c^2, and e^2 reflect the total additive genetic, shared environmental, and non-shared environmental variance (e.g. common and specific combined). Standardized variance estimates may not add up to 1.00 due to rounding.

Discussion

The current study utilized a longitudinal adoption design to examine the magnitude and developmental patterns of genetic and environmental influences on substance use from ages 14-

18. Importantly, results from adoption studies can be used to anchor estimates of environmental influences which are indirectly assessed from twin studies, but directly estimated in sibling adoption designs.

Due to limited sample sizes, multivariate analysis of the use/no use data was not feasible. Age specific univariate analyses of substance use/no use at each age yielded parameter estimates with large confidence intervals. However, the point estimates suggested interesting trends. In contrast to many twin studies, which tend to show more evidence of environmental influences during the adolescent years (Rose et al., 2001), the pattern of sibling pair tetrachoric correlations from age 14-18 indicates moderate heritabilites for liability to use cigarettes, alcohol and marijuana in adolescence. Heritability decreased in magnitude for cigarette and alcohol use across adolescence, but increased for marijuana use. Shared environmental influences were relatively modest for cigarette use/no use across adolescence. For alcohol use, there is a trend for increasing shared environmental influences with the greatest influence at age 18, where access to substances may be more readily available. In comparison, a recent longitudinal adoptive parent-offspring study found significant shared environmental (parent-offspring) influences on drinking behavior at this age, while genetic influences were important in early adulthood (McGue et al., 2014). A twin study by Kendler et al. (2008) also found that shared environmental influences on liability to use alcohol remain well into the young adult years. In contrast, Koopmans et al. (1997) found substantial early (age 12-14) shared environmental influences for males only, while female alcohol use had strong early genetic influences. In contrast to Kendler et al. (2008), our study found that shared environmental influences on liability to use marijuana were modest across the range from age 14 to age 18. Similarly, Baker et al. (2011) described a common factor

model with substantial genetic effects on marijuana and illicit drug use/no use at age 13-14, with few additional innovative genetic affects emerging at ages 16-17 and 19-20.

Our adoptive and control sibling correlations for quantity/frequency of substance use generally suggest genetic influences, with only modest effects of the shared environment, particularly at early ages when prevalence of use was lower. Additive genetic factors have also been shown to contribute substantially to substance use across development. A meta-analysis by Bergen, Gardner, & Kendler (2007) found an increase in the heritability for multiple phenotypes from adolescence into adulthood but no significant increases for two substance use measures (i.e. nicotine initiation and alcohol consumption).

Although substance use is correlated across measurement occasions, it is possible that particular environmental shifts (e.g. starting high-school) or biological changes (e.g. beginning puberty) may influence behavior at specific periods of adolescence. Thus, we fitted multivariate biometric models to test whether use patterns across five ages had common influences or age-specific influences.

Overall, all age-specific genetic and shared environmental influences could be dropped from cigarette, alcohol, and marijuana quantity/frequency of use models (e.g. Models 3-5 in 2.6). Age-specific non-shared environmental influences may reflect measurement error rather than unique environmental influences that could influence substance use at multiple waves.

While most variance was due to common influences, it is possible that common factors could have varying degrees of influence over adolescence. We tested this by constraining loadings from common factors to be equal across ages (Models 6-9). There were some non-significant trends for common additive genetic influences, in that the proportion of variance explained for cigarette and alcohol quantity/frequency of use appeared to increase as participants

aged (p=.06, .09, respectively). For marijuana, these influences were the largest at ages 15 and 16, though path loadings could be constrained across ages without significant decrement in fit compared to the base independent pathway model (p=.07). Common shared environmental pathways were stable across ages for all substances (p=.24 - .66). Common non-shared environmental pathways were highly variable and could not be constrained for any substance. Given that few age-specific influences were detected, the total proportion of variance explained by additive genetics and shared environment follow similar trends.

There are several limitations to consider when interpreting these results. A potential confound of the CAP sample is that there are more same-sex sibling pairs in the control families, while the adoptive families include more opposite-sex pairs. If same-sex sibling pairs are more similar than opposite-sex pairs on substance use behaviors, the increased similarity of the control families (due to greater numbers of same-sex siblings) could bias our estimates of variance due to genetic effects upward. To test this, we ran a series of regression analyses to test the effect of adoption vs. control status, same sex vs. opposite sex status, and their interaction on sibling pair difference scores for quantity/frequency of use. Across five time points for each substance, same sex pairs were not significantly more similar than opposite sex pairs nor were these effects different across adoptive and control families. For use/no use, we used logistic regression to test the same effects on pair concordance and discordance. Across five time points for each substance, the test of the same sex/opposite sex effect was significant only once. However, the effect was in the opposite direction than expected. Opposite sex pairs were more similar for age 17 alcohol use than same sex pairs, and this was more true for adoptive pairs than control pairs. Thus, there is no evidence in our data to suggest that the greater similarity of control siblings

compared to adoptive siblings can be explained by the difference in same-sex versus opposite-sex pairs.

Second, we did not have identical assessment questions throughout the length of the study. Our transformation from 6 month use or past year use variables into past month variables required some assumptions; namely, that average use over the past month was consistent with the given time span. For example, if a participant reported using marijuana once a month on average over the past six months (or year), they would have been coded as using once during the past month although they may have used more or less during different peak times over the year.

Finally, the numbers of adoptive and nonadoptive sibling pairs available at each age were relatively small in this study. This was primarily due to the requirement that both proband and sibling be tested within the same test age year—which was necessary for yearly assessment of the sibling pairs. This lead to some sparse data issues that limited our approaches to data analysis (e.g. multivariate analysis of the use/no use was not possible; and multivariate analysis of the quantity/frequency data required use of methods that were robust to sparse data issues).

Despite these limitations, our study provides a unique contribution to the literature on genetic and environmental influences on substance use behavior. As the first sibling-based longitudinal adoption study of substance use, our estimates provide a test of the role of environment on use of cigarettes, alcohol, and marijuana from adolescence into early adulthood. These estimates corroborate the point estimates of cross-sectional twin studies and other prospective designs. Importantly, the general trend of increasing genetic influences in late adolescence/early adulthood for quantity/frequency of alcohol use mirrors results reported from a recent parent-offspring longitudinal adoptive design (McGue et al., 2014). In conclusion, results of the present study indicate that individual differences in substance use from 14 to 18 years of

age are largely due to common influences. Moreover, although the sample of adopted and control sibling pairs was relatively small, our findings suggest that frequency/quantity of substance use during adolescence are due substantially to genetic influences, and that new genetic influences may emerge for cigarette and alcohol use in late adolescence.

CHAPTER III

ETIOLOGICAL OVERLAP BETWEEN SEX UNDER THE INFLUENCE AND NUMBER OF LIFETIME SEXUAL PARTNERS

Background

Risky sexual behaviors (RSB) are those that increase one's risk for contracting human immunodeficiency virus (HIV) and other sexually transmitted infections. Within the United States, there are about 19.7 million new cases of sexually transmitted infections each year (Satterwhite et al., 2013), which translates to an estimated $15.6 billion dollars of direct treatment costs (Owusu-Edusei et al., 2013). As a result, there has been widespread public interest in reducing RSB and related outcomes: from sexual education in public schools, to "safe-sex" campus initiatives, and community based preventative healthcare programs that aim to reduce infections and unintended pregnancies. Many of these programs focus on the link between substance use and sexual risk, with the hope that reducing drug and alcohol induced impairment will lead to safer sexual practices.

While drug and alcohol use is associated with higher rates of RSB on a population level, these behaviors could be correlated for several reasons: 1) drug and alcohol use during sexual encounters may cause people to take more risks or 2) there are third variables (e.g. environmental or genetic factors) that lead to both substance use and RSB. In terms of public health, identifying the optimal approach to reduce HIV, sexually transmitted infections, and other maladaptive outcomes relies on understanding the causal structure of these related behaviors. For instance, substance use prevention may be useful in its own right, but it may have little effect on overall RSB related outcomes if the association is due to other confounding factors such as parental monitoring or a proclivity towards impulsive behavior.

While there is a wealth of literature focused on the relationship between alcohol (and to a lesser extent other drug use) on RSB, the evidence is mixed (Leigh & Stall, 2008). In a meta-analysis of 12 randomized control studies, exposure to alcohol had a direct effect on RSB intent (e.g. one's likelihood to have unprotected sex; Rehm, Shield, Joharchi, & Shuper, 2012). However, an additional meta-analysis could not rule out alternative (i.e. non-causal) explanations for the link between alcohol use and HIV incidence, which is arguably a more proximal measure to actual RSB than RSB intent (Shuper, Neuman, Kanteres, Baliunas, Joharchi, & Rhem, 2010). Additionally, there is some event-based evidence that real-world behaviors do not change due to alcohol exposure. Telephone and diary tracking studies found that condom use patterns were similar across sober encounters compared to those where alcohol had been consumed (Morrison, Gillmore, Hoppe, Gaylord, Leigh, & Rainey, 2008; Leigh, Vanslyske, Hoppe, Rainey, Morrison, & Gillmore, 2008) or when marijuana had been used (Walsh, Fielder, Carey, & Carey, 2013), although some interactions with partner type (e.g. causal vs. regular) were reported. It is possible that both explanations are at least partially true, in that there is both a direct effect of substance use on RSB and there are underlying influences on both. Using a quasi-experimental design in which many genetic and environmental confounds are shared between twin pairs (McGue, Osler, & Christensen, 2010), biometrical modeling is a useful tool for exploring the *extent to which* either of these explanations are supported. By comparing cross-twin cross-trait correlations, the likelihood of the following scenarios can be estimated: 1) drug and alcohol use during sex and RSB share the same influences (i.e., correlated liabilities model), 2) drug and alcohol use during sex *causes* higher RSB, and 3) there is a pattern of reverse causation (Neale & Kendler, 1995; Rhee et al., 2005).

Bivariate twin modeling is used to estimate how much of this association is due to genetic or environmental sources (i.e., a re-parameterization of the correlated liabilities model). These models allow for calculation of the additive genetic correlation (rA or rG), which measures the extent to which the additive genetic influences (additive effects of all contributing loci) on one variable are shared with the additive genetic influences of the second variable. Shared environmental correlations (rC) and non-shared environmental correlations (rE) can also be computed, with values ranging from -1 to 1. While a significant rA is indicative of pleiotropy (i.e. when genetic variants influence more than one trait); the bivariate twin model cannot distinguish between the following types: 1) biological pleiotropy (e.g., when the same genetic factors have a direct biological influence on both traits independently), 2) mediated pleiotropy (e.g., when one trait is causally related to a second trait so that the genetic factors for the first trait are indirectly associated with the second), and 3) spurious pleiotropy (e.g., the appearance of pleiotropy due to misclassification or ascertainment bias; Solovieff et al., 2013). Similarly, environmental influences will be associated with both traits if: 1) there are direct influences on both traits, or 2) if the first trait causes the second trait, those influences on the first trait in turn will influence the second trait indirectly. Evidence for correlated liabilities is necessary but not sufficient for causal inference.

Different assumptions about causality can be made under the following scenarios. 1) Under the assumption that sex under the influence of no familial influences on sex under the influence, covariance should be explained entirely by nonshared environmental influences. In this case, the possibility of biological pleiotropy or shared environmental third variable influences can be ruled out. However, this pattern could be indicative of causality or nonshared direct influences on both traits. Similarly, this pattern is expected only if the first trait is a

"random exposure," which is unlikely for most behavioral traits (e.g., rMZ would be equal to rDZ). 2). Non-causality may be assumed if all covariance is due to common genetic and shared environmental sources and non-shared environmental covariance is non-significant (e.g., suggesting that mediated pleiotropy or mediated shared environmental influences are unlikely). In the case of causality, any significant non-shared environmental effects on the first trait should also explain covariance with the second trait. In the final case, 3) covariance could be significantly explained by all three sources. In this case, direction of causation models (DoC) models (Heath et al., 1993.) are required to determine the likelihood of causation.

DoC models test assumptions about causality for traits measured in genetically informed samples, given that certain conditions are met. In order to distinguish between the two causal models (drug and alcohol use during sex causes an increase in RSB, compared to the reverse), the two traits must have different modes of inheritance (the magnitude of genetic vs. environmental influences must vary between traits). While tests of reciprocal causation (both directly cause each other over time) are theoretically possible, large samples and multiple trait indicators are required (Neale, Duffy, & Martin, 1994). Additionally, it is assumed that measurement error for the two traits is uncorrelated between relatives.

To our knowledge, this is the first twin or family study directly exploring the nature of the relationship between sex under the influence and a measure of RSB in early adulthood. The current study investigates the role of drug and alcohol use on RSB using several self-report measures. First, we aimed to replicate the phenotypic relationship between drug and alcohol use during sexual decision making (a composite measure hereinafter referred to as sex under the influence) and RSB—assessed by number of lifetime sexual partners (corrected for age and sex). Lifetime number of sexual partners has been shown to be a robust measure of RSB, and is a

strong predictor of sexually transmitted infections in several samples (Karlsson, et al., 1995; Santelli et al., 1998; Sturdevant et al., 2001; Epstein et al., 2013).

Using genetically informed twin designs (Neale & Cardon, 1990), we used structural equation modeling to partition the covariance between these variables into genetic and environmental components. Model results where all covariance was explained by non-shared environmental sources would be most consistent with a causal model. Common genetic and shared environmental sources of covariance *may* be inconsistent with causality. Alternatively, it is possible that genetic and shared environmental influences on the liability to engage in sex under the influence also influence RSB through a direct pathway. Thus, DoC models were used to test the likelihood that sex under the influence causes an increase in lifetime number of partners, as well as test the possibility of reverse causation. Finally, we hypothesize that RSB may be correlated with drug and alcohol use both within and outside of sexual contexts, thus we test how controlling for general substance use mediates the relationship between sex under the influence and number of lifetime sexual partners. We model whether general drug use can mediate this relationship genetically or environmentally.

Methods

Sample

Participants were drawn from the third wave of data collection from the Center on Antisocial Drug Dependence; PI: J. K. Hewitt), a longitudinal study of adolescent/young adult antisocial behavior and substance use, which includes four genetically-informative samples. The Colorado Longitudinal Twin Study and the Colorado Community Twin Study were included in the primary analysis. Twins in the analysis were in early adulthood (female m=25.24 years, *s.d.*= 2.50, n=1047; male m=25.18, *s.d.*= 2.66, n=823) and were representative of Colorado

demographics (Rhea et al., 2006; 2013). While the additional two CADD samples were not included in the primary analysis, all four of the community-based Center samples were used to calculate age- and sex-normed measures of substance use and RSB. The additional samples included the Colorado Adoption Project which follows adoptive children, matched controls, and their families (Petrill et al., 2003), and the control participants of the Colorado Family Study which is comprised of probands formerly in treatment for adolescent antisocial drug dependence, their siblings, and matched control families (Stallings et al., 2003).

Measures

Sexual behavior was assessed using the Modified Risk Behavior Questionnaire (M-RBQ, adapted from Booth, Corsi, & Mikulich-Gilbertson, 2004). Due to the sensitive nature of the items on this questionnaire, the M-RBQ was administered via a computer program in a private room and a "would rather not answer" option was available for all items. While this did result in some missing data, this option was seldom chosen and did not seem to be correlated with other attributes of the participant. *Number of lifetime sexual partners* was scored from a single item (i.e. "in your lifetime, with how many people (different partners) have you had oral, vaginal or anal sex?"). Scores were quasi-continuous and measured on a seven point scale ("none" [0], "one" [1], "two" [2], "three-five" [3], "six-nine" [4], "ten-nineteen" [5], and "twenty or more" [6]). Scores were corrected for age using standard regression procedures and then z-scored within sex.

Sex under the influence was a composite of four items (adapted from items used by the Minnesota Center for Twin and Family Research; Iacono, McGue, & Krueger, 2006), which assessed the extent to which drug and alcohol use co-occurred with sexual decision making. Items assessed the frequency in the past 12 months that participants 1) had alcohol or drug use

influence a decision to do something sexual with a partner, 2) did more sexually with a partner than planned due to drinking or drug use, 3) used drugs or alcohol to feel more comfortable with a sexual partner, or 4) had unprotected sex due to drinking or drug use. Each item was assessed on a seven-point scale ("never" [0], "one time" [1], "two times" [2], "three-five times" [3], "six-nine times" [4], "ten-nineteen times" [5], or "twenty or more times" [6]). Participants who either had 1) no sexual partners in their lifetime, or 2) only reported one partner within the past 5 years were not assessed on these questions (they skipped out of this assessment), but were coded zero since their drug and alcohol use would have little to no effect on sexual risk taking (Derks, Dolan, & Boomsma, 2004). Additionally, participants who answered "never" on all four individual items were scored zero and were presumed to be low in sex under the influence and general sexual risk taking. Remaining participants were split into quartiles based on their composite scores to form a 0 to 4 ordinal index of sex under the influence, with the 4^{th} quartile indicating the highest risk.

An index of *general substance use* was created from several measures on the Composite International Diagnostic Interview- Substance Abuse Module (CIDI-SAM; Robins et al., 1988) as a way to capture general participant endorsed substance use. Participants were considered 'users' of a drug class (e.g., marijuana, stimulants, sedatives, club drugs, cocaine, heroin /opioids, PCP, hallucinogens, or inhalants) if they reported using a class of drugs more than 5 times in their life. However, participants were considered to be alcohol users if they had ever had more than one drink. These thresholds were determined by the administration algorithms of the CIDI-SAM. Subjects skipped out of drug assessment categories if they did not meet these minimum use thresholds, so sub-threshold use is not recorded by the CIDI-SAM. Since tobacco is not typically thought to play a role in sexual decision making, it was not included in the final

computed variable. Given the relative scarcity of scores at the high range of use, higher scores were truncated resulting in a six point ordinal scale ("not a user of any substances" [0], "one substance" [1], "two substances" [2], "three substances" [3], "four substances" [4], "five or more substances" [5]; Derks, Dolan, & Boomsma, 2004).

Zygosity. For 92% of the subjects DNA was extracted from buccal cells through saliva and/or cheek swabs, and zygosity was confirmed by analyzing 11 highly polymorphic short tandem repeat (STR) polymorphisms (Smolen, 2005). For twins yet to provide DNA samples, zygosity was established via repeated tester ratings on a 9-question survey (Nichols & Bilbro, 1966; see Rhea et al., 2013 for details).

Analyses

All descriptives and composite variables were computed using SPSS version 21 (SPSS, 2012). Phenotypic correlations, twin correlations, and structural equation models were estimated using Mplus version 7.4 (Muthén & Muthén, 1998-2012) using raw data options and employing maximum likelihood estimation (ML) for univariate analyses of the continuous variable and robust weighted least squares (WLSMV) in the case of ordinal variables and multivariate analyses.

Model fit was determined using the following indices: root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Akaike information criterion (AIC), the latter of which could only be estimated for continuous variables. Nested models were compared using likelihood-ratio χ^2 difference tests using the difference in degrees of freedom (*df*) in the full model compared to more constrained models.

Nested models tested 1) significance of sex differences in means/thresholds and path estimates, and 2) significance of each pathway (e.g. comparing full model to a model dropping

the path of interest). The DIFFTEST in Mplus was used in model comparisons using WLSMV estimation to address biases in the χ^2 test produced by ML estimation for ordinal variables. In rare cases of non-convergence, the significance level of specific path estimates was determined by testing whether the Mplus-computed confidence interval included 0.

Biometrical Modeling

Univariate structural equation models were fit to data for each the three variables. Using a genetically informative twin sample, these models are able to partition the sources of variance of these variables into additive genetic, or sum effect of each individual segregating allele (a^2), interactive effects of alleles within loci (referred to as dominance effects; d^2), shared or common environmental (c^2), and non-shared environmental (e^2) factors (See Figure 1.1). Estimates are derived from the difference in MZ twin similarity, who share 50% of their independently segregating alleles (e.g., 100% of additive and dominance effects) and DZ twin similarity, who share 50% of their segregating alleles on average (e.g., 50% of additive effects and 25% of dominance effects).

A bivariate Cholesky decomposition of covariance (e.g., a re-parameterization of the correlated liabilities model) was used to explore the nature of the relationship between sex under the influence and number of lifetime sexual partners, which follows a similar logic. Much like variance of a single trait, the covariance between traits can be decomposed into genetic and environmental sources. To the extent that additive genetic effects explain the co-occurrence of these behaviors, we would expect that MZ cross-twin cross-trait correlations to be higher than DZ correlations (e.g., a twin's sex under the influence score should be more strongly associated with their co-twin's number of lifetime sexual partners within MZ pairs compared to DZ pairs if there are common additive genetic factors that contribute to both variables). Although bivariate

twin models were used to obtain parameter estimates, the magnitude of additive genetic covariance can be estimated from the difference between MZ and DZ pair cross-twin cross-trait correlations. Thus, if cross-twin cross-trait correlations were not significantly different between MZ and DZ pairs we would assume all covariance is due to shared environment. Finally, non-shared environmental factors will capture the covariance between the variables within individuals (e.g., in a case where all covariance is explained by E, there would be no cross-twin cross-trait resemblance for either MZ or DZ twin pairs). As such, the Cholesky model decomposes sources of variance and covariance into multiple A, C (or D), and E factors. In the bivariate Cholesky, the first A, C (or D), and E factors model the covariance between the variables. The secondary A, C (or D), and E factors are orthogonal to the first factors and model the residual variance in the second variable. A trivariate Cholesky factorization is a straightforward generalization of the bivariate model: the first factors of a trivariate model load on all variables to model shared covariance. The secondary factors capture only residual variance and covariance between the second and third variables, while the final factor captures the unexplained variance of the final variable.

A trivariate Cholesky decomposition model was used to explore the nature of the covariance between sex under the influence and number of lifetime sexual partners, when controlling for general substance use. As a result, this model directly tests whether general propensity towards substance use is a direct mediating variable and if so, whether this mediation is of genetic or environmental origin.

The DoC models are nested under the bivariate model. Rather than allowing A, C (or D), and E cross paths to model the covariance between sex under the influence and number of lifetime sexual partners, all covariance is modeled in a single direct pathway. Given causality,

the magnitude of each variance component (i.e., A, C [or D], and E) should be approximately equal in magnitude. Thus, a single parameter represents the magnitude of the A, C (or D), and E factors which directly influence sex under the influence and have an indirect (mediated) influence on number of lifetime sexual partners. For data inconsistent with causality, constraining these influences on lifetime number of partners to be fit through a single direct pathway from impaired influences should significantly reduce the model fit. To rule out reverse causation, we also fit a model with a direct pathway from number of lifetime sexual partners to sex under the influence.

To summarize how the models inform the research question, the covariance structure across these traits will either be inconsistent with causality or inconclusive (e.g., causality cannot be ruled out). Results with significant cross path loadings from the E factor with no genetic and shared environmental covariance would be consistent with a causal model or evidence of an environmental influence (e.g. third variable influencing both traits) that is not shared across twins. Models that are inconsistent with causality would result in a non-significant path from the E factor, but significant A, C (or both) cross paths. For instance, a significant cross path from the A factor could indicate either biological or mediated pleiotropy; however, if the overlap was solely due to mediation the E path would also be expected to be significant. If the E cross paths and A, C (or both) cross paths are found to be significant, DoC models are required to draw inferences about causality. The DoC models are nested under the bivariate Cholesky. If the forced constraints result in a significant reduction of model fit, a causal model is unlikely. Finally, we examine the extent to which propensity toward substance use may explain the overlap between sex under the influence and number of lifetime sexual partners in general. The trivariate model shows how the magnitude of the genetic, shared environmental, and nonshared

environmental overlap between sex under the influence and number of lifetime sexual partners may change when controlling for general substance use.

Results

Males reported slightly higher mean scores compared to females on all variables (See Table 1.1). The modal number of lifetime sex partners was 3 to 5 for both males and females in this age group, but nearly twice as many males reported 20+ partners (10.4%) compared to females (5.4%). Across all twins included in the analysis, sex under the influence was significantly associated with number of lifetime sexual partners (r=.600). Additionally, our measure of general substance use was substantially correlated with both sex under the influence (r=.565) and number of lifetime sexual partners (r=.515). For phenotypic correlations computed separately by gender, see Table 3.2. Twin correlations for the three variables (see Table 3.3) were suggestive of genetic influences on all three variables, as MZ correlations were consistently higher than DZ correlations. We determined ACE models were more appropriate than ADE models, which was consistent with patterns identified in review of substance use behavior (Sullivan & Kendler, 1999; Stallings, Gizer, & Young-Wolff, 2016) and risky sexual behavior (Harden 2014). Given differences in male and female correlations, sex limitation models were tested (i.e., formal tests of gender differences were conducted).

Table 3.1

Frequency of raw scores, by gender

	Females	Males
1) General substance use		
None	3.8% (n=40)	4.5% (n=37)
1	57.0% (600)	41.2% (339)
2	27.2% (286)	31.1% (256)
3	5.2% (55)	9.1% (75)
4	3.1% (33)	6.2% (51)
5-10	3.6% (38)	7.9% (65)
2) Sex under the influence		
0	65.6% (n=669)	60.5% (491)
1	7.1% (72)	11.7% (95)
2	10.0% (108)	8.1% (66)
3	21.7% (97)	8.8% (71)
4	7.8% (80)	10.9% (88)
3) Number of lifetime sexual partners		
None	8.6% (n=88)	9.0% (74)
1	16.5% (173)	14.5% (118)
2	11.0% (115)	8.4% (69)
3-5	21.7% (227)	22.2% (183)
6-9	19.0% (199)	16.9% (139)
10-19	15.5% (162)	17.7% (146)
20+	5.4% (57)	10.4% (86)

Table 3.2

Phenotypic correlations, by gender

	General substance use	*Sex under the influence*	*Number of lifetime sexual partners*
1) General substance use	1	**.617**	**.487**
2) Sex under the influence	.505	1	**.606**
3) Number of lifetime sexual partners	.561	.597	1

Note. Female estimates listed on lower diagonal, male estimates listed on upper diagonal in **bold**.

Table 3.3

Twin correlations for the three variables by type of twin

	MZ-F	MZ-M	DZ-F	DZ-M	DZ-OS
1) General substance use	.644	.760	.372	.573	.478
2) Sex under the influence	.347	.345	.334	.212	.193
3) Number of lifetime sexual partners	.661	.655	.335	.365	.312
(number of pairs)	254-263	171-176	147-155	129	135-140

Note. F= female, M=male, OS=opposite sex. Polychoric correlations were estimated in MPlus.

Univariate Model: Test of sex limitation

We first tested whether mean/threshold gender differences should be accounted for in our models. For each variable, we tested whether thresholds for categorical variables and means for continuous variables could be constrained to be equal across gender in models that also estimated separate path coefficients for males and females. Male and female distributions for general substance use were significantly different, so that thresholds could not be constrained across gender ($\chi^2_{diff}(5)= 62.29$, $p< .001$). Alternatively, thresholds were not significantly different across gender for sex under the influence ($\chi^2_{diff}(4)= 8.292$, $p=0.081$). We expected no mean differences across gender for number of lifetime partners as it was z-scored within gender, and this constraint did not result in a decrement of fit ($\chi^2_{diff}(1)= 1.37$, $p=0.242$). Thus, separate thresholds for males and females were used in subsequent models for *general substance use,* while means and thresholds were constrained across gender for sex under the influence and number of lifetime sexual partners.

Second, we conducted a quantitative sex limitation model that tested whether the A, C, and E factor loadings could be constrained across gender. Constraining these estimates, when there are significant differences in the magnitude of these influences between males and females, will result in significant χ^2 difference tests in nested model comparisons. While there were no quantitative gender differences in the biometrical factor loadings for sex under the influence

($\chi^2_{diff}(3)$= 5.22, p=0.157) and number of lifetime sexual partners ($\chi^2_{diff}(3)$= 2.24, p=0.523), there was a significant decrement in model fit in the constrained model for general drug use ($\chi^2_{diff}(3)$= 8.05, p=.0449). For drug use genetic influences were significantly greater for females than for males while shared environment made a greater contribution to variation in drug use for males.

Influences of the A, C, and E factors are reported as standardized path estimates (i.e. a_{11}, c_{11}, and e_{11}; reported in Table 3.4). When squared, path estimates represent the proportion of variance explained by a specific influence (e.g. a_{11}^2 is equivalent to the proportion of variance explained by additive genetic variation, also termed heritability; see Figure 1.1). Heritability was greatest for number of lifetime sexual partners (a_{11}^2= $.710^2$ = 50%) and general drug use ($a_{11}^2{}_{male}$= $.612^2$ =37%, $a_{11}^2{}_{female}$= $.700^2$ = 49%), while sex under the influence were moderately heritable (a_{11}^2= $.434^2$ = 19%). The greatest evidence of shared environmental influences on a single trait was for general drug use, specifically for males ($c_{11}^2{}_{male}$= $.622^2$=37% of variance explained by shared environment, $c_{11female}^2$= $.389^2$= 15% of variance explained by shared environment).

Table 3.4.

Standardized path loadings for univariate models

Model		a_{11}^2	c_{11}^2	e_{11}^2
1) General substance use	**Female**	**0.49***	**0.15***	**0.36***
	Male	**0.37***	**0.39***	**0.24***
	Constrained	0.46*	0.24*	0.30*
2) Sex under the influence	Female	0.02	0.32	0.65*
	Male	0.27	0.07	0.65*
	Constrained	**0.19**	**0.16**	**0.65***
3) Number of lifetime sexual partners	Female	0.42*	0.15	0.44*
	Male	0.61*	0.04	0.35*
	Constrained	**0.50***	**0.09***	**0.40***

Note. Female and male paths estimated separately are compared to a constrained model with paths estimated jointly. **Bold** indicates the best fitting model; fit statistics: 1) $\chi^2 = 42.335$, df = 49, p=0.7384, RMSEA < 0.001, CFI = 1.000; 2) $\chi^2 = 47.446$, df = 42, p=0.2603, RMSEA = 0.025, CFI = 0.874; and 3) $\chi^2 = 31.198$, df = 21, p=0.0704, RMSEA = 0.049, CFI = 0.959, AIC= 4975.576. Estimates do not add to 1 due to rounding.

Bivariate Model: Covariance between sex under the influence and number of lifetime sexual partners

As expected from the lack of gender differences in the univariate models for sex under the influence and number of lifetime sexual partners, constraining path estimates across gender in the bivariate model did not significantly reduce the fit of the model ($\chi^2_{diff}(9)= 10.425$ p=0.317). Though the constrained bivariate model had the best fit, we report paths separately for the sake of completeness as gender differences emerged in the trivariate model.

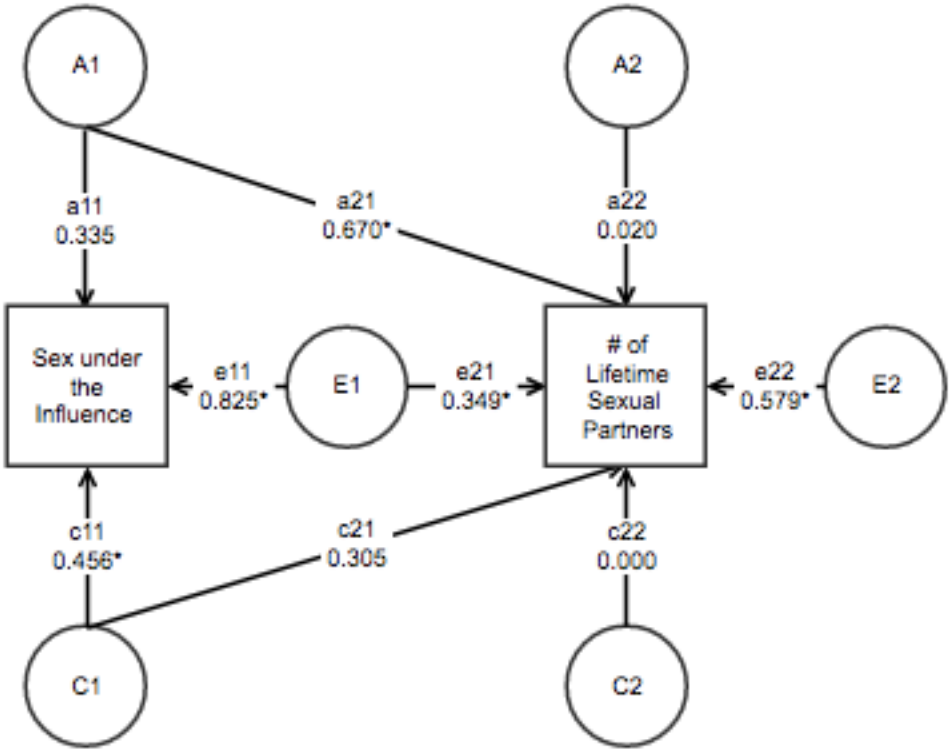
Significance of the cross paths (a_{21}, c_{21}, and e_{21}) are dependent on the size of the loading on the first factor (a_{11}, c_{11}, and e_{11}, respectively). For example, only a small proportion of the covariance would be explained by additive genetics if the first variable had a small heritability- regardless of the magnitude of the cross path. Thus, significant cross paths are of interest because they indicate a source of third- variable confounding between sex under the influence and number of lifetime sexual partners. Finally, loadings on the final factor represent unique

influences not shared with the first variable (e.g. genetic or environmental factors that influence number of lifetime sexual partners but not sex under the influence).

In the best fitting model (constrained across gender), both genetic and environmental factors contribute to the correlations between sex under the influence and lifetime number of sexual partners. Figure 3.1 shows there was significant additive genetic (a_{21}=0.670, p= .0164) and nonshared environmental (e_{21}=0.349, p <.001) covariance or overlap. Though the shared environmental influences on number of lifetime sexual partners were shared entirely with those on sex under the influence (rC=1.00), the parameter of interest was not significant (c_{21}, p=.0744).The proportion of covariance explained by these factors was as follows: A= 57.7%, C= 26.7%, and E= 15.6%, suggesting that much of the covariance between sex under the influence and number of lifetime partners was due to common genetic factors.

Figure 3.1

Bivariate model with standardized path loadings constrained across gender



Note: Model fit statistics: $\chi^2 = 91.590$, $df = 80$, $p = 0.1768$, RMSEA = 0.027, CFI = 0.979. Pathway significance determined through model comparison, significant pathway denoted with * ($p<.05$).

However, examination of the gender-specific estimates suggests that genetic factors contribute more to this correlation for males, while environmental factors contribute more in females. In females, the parameter signifying genetic overlap in the unconstrained model is not significant (a_{21}= 0.719, p=0.1808). Since sex under the influence was less heritable in females (a^2 = 0.02), compared to males (a^2 = 0.27), genetic overlap explains a smaller proportion of the total covariance between sex under the influence and number of lifetime sexual partners in females—despite the larger point estimate for parameter a_{21}. In contrast, shared environmental influences appear to explain a higher proportion of variance for sex under the influence in females; however the specific parameter signifying shared environmental overlap did not reach significance

(c_{21}=0.335, p=0.1572). Though path estimates could be constrained, separate path estimates are reported in Table 3.5 for sake of completeness, as gender effects emerge in the trivariate model.

It is important to point out that the genetic covariance (as well as the environmental covariance) is the product of path coefficients, in this case a_{11} x a_{21}. Although the estimate of a_{21} in females is greater than that for males, the product of the two path coefficients is greater for males than for females. As power to detect significance was reduced by using ordinal variables, it may be useful to examine the estimated proportion of covariance explained by each factor (e.g., for females and males, respectively, A= 55.8% and 67.7%, C= 31.7% and 11.5%, and E= 12.4% & 20.8%).

Table 3.5

Bivariate model standardized path loadings by gender

	a11	a21	a22	c11	c21	c22	e11	e21	e22
Female	0.249	0.719	0.000	0.542	0.335	0.007	0.803*	0.339*	0.505*
Male	0.482*	0.695*	0.001	0.287	0.067	0.019	0.882*	0.385*	0.603*

Note. Model depicted for one twin only**.**

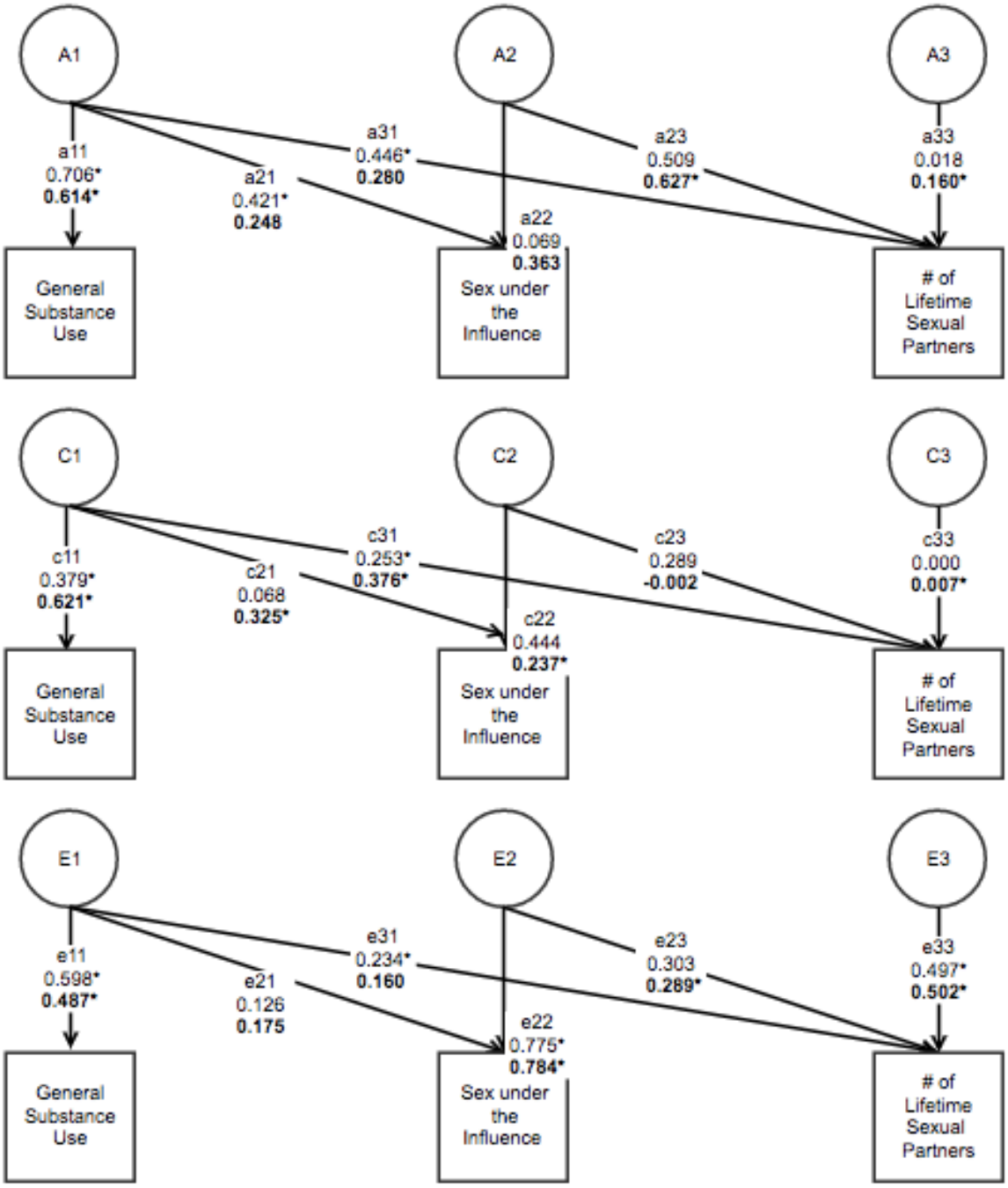
Trivariate Model: General substance use as a mediator

The trivariate model allows for the decomposition of the residual covariance between sex under the influence and number of lifetime sexual partners after accounting for the common covariance shared between all three variables (See Figure 3.2). That is, this model allowed us to test whether sex under the influence or impaired sexual decision making due to drugs and alcohol remained an important predictor or correlate of risky sexual behavior, after controlling for drug use in general. Again, we tested the significance of the parameters representing covariance between sex under the influence and number of lifetime sexual partners (i.e., a_{23}, c_{23}, & e_{23}, or the cross paths between the second and third latent factors), as well as the significance of covariance pathways (a_{12}, c_{12}, and e_{12}) between general substance use and sex under the influence.

We report the constrained model for consistency with the reports of the univariate and bivariate model (See Table 5.6); however, evidence of sex differences emerged when trying to constrain male and female paths to be the same in the trivariate model ($\chi^2_{diff}(18)= 44.349$, $p<.001$). To appropriately interpret the model of best fit, males and female patterns should be considered separately.

Figure 3.2

Trivariate model with standardized path loadings estimated for each gender



Note. Female path estimates listed above, male path estimates are listed **bold**. Model fit statistics: $\chi^2 = 157.169$, $df = 148$, $p = 0.2874$, RMSEA = 0.018, CFI = 0.995.

Table 3.6

Trivariate model with standardized path loadings constrained across gender

	11	21	31	22	23	33
a	0.678*	0.307*	0.488*	0.267	0.482	0.282
c	0.491*	0.247	0.186	0.341	0.200	0.001
e	0.547*	0.162*	0.179*	0.794*	0.289*	0.510*

Note. Model depicted for one twin only.

The bivariate results suggested that common genetic factors make an important contribution to the covariance between sex under the influence and number of lifetime partners—particularly for males. After controlling for substance use in general, the genetic covariance between sex under the influence and number of lifetime partners is reduced in males but remains significant. That is, although part of the genetic covariance is mediated through genetic factors in common with substance use in general (a_{31}), other distinct genetic influences contribute to the covariance between sex under the influence and number of lifetime partners (a_{23}). However, nonshared environmental influences specific to sex under the influence and number of lifetime partners (e23) still remain for males when substance use in general is included; thus, causality cannot be ruled out with these models alone.

For females, after controlling for substance use in general, there was no significant covariance specific to sex under the influence and number of lifetime partners. Though the first C and E factors describe covariance shared among all three variables, the path loadings onto sex under the influence were not significant (c_{21}, p= .6930, & e_{21}, p=.1184). This pattern suggests that rather than a causal relationship, much of the relationship between sex under the influence and number of lifetime partners was mediated through influences in common with drug use in general (a31, c31, e31).

It was possible that males and females differ not only in the extent to which genetic or environmental factors influence a trait (i.e., quantitative test discussed above), but that there are

different genetic or environmental influences across gender (qualitative sex differences). Qualitative sex difference could occur even if the genetic influences were the same magnitude across genders (e.g. equal hertiabilites across males and females, albeit from different genes). Inclusion of DZ-OS (e.g. opposite sex dizygotic) pairs allows for such tests. If different genes influence female versus male behavior, we would expect that the rDZ-OS (opposite sex twin correlation) would be lower than rDZ-SS (same sex twin correlation). For our three variables, rDZ-OS were not significantly less than rDZ-SS (see Table 3.3) suggesting that the genes that influences male traits are likely the same genes that influence female traits. Inclusion of DZ-OS twins changes the interpretation of the significance of given path estimates. Notably, when testing the significance of female or male specific paths via model comparisons (e.g. comparing a full model to a model dropping the path of interest), we are essentially fixing both the path to be zero and a portion of the covariance between males and female pairs to be zero. That is, dropping a female (or male) parameter would not only test the significance of a female (male) loading but also the common pathways across males and females). We re-ran all of the models excluding DZ-OS pairs and estimated paths were of similar magnitude to those reported here, although some significant paths dropped to non-significance due to a reduced sample size (see Tables 3.7-3.9).

Table 3.7

Standardized path loadings for univariate models excluding OS twin pairs

Model		a_{11}^2	c_{11}^2	e_{11}^2
1) General substance use	**Female**	**0.54***	**0.10**	**0.36***
	Male	**0.37***	**0.39***	**0.24***
	Constrained	0.46*	0.24*	0.30*
2) Sex under the influence	Female	0.02	0.32	0.65*
	Male	0.27	0.07	0.65*
	Constrained	**0.13**	**0.21**	**0.65***
3) Number of lifetime sexual partners	Female	0.42*	0.15*	0.44*
	Male	0.53*	0.11	0.36*
	Constrained	**0.46***	**0.13***	**0.40***

Note. Female and male paths estimated separately compared to a model constrained with paths estimated jointly. **Bold** indicates the best fitting model; fit statistics: 1) $\chi^2 = 43.208$, $df = 41$, $p = 0.3772$, RMSEA = 0.016, CFI = 0.998, 2) $\chi^2 = 40.637$, $df = 33$, $p = 0.1694$, RMSEA = 0.034, CFI =0.814 , 3) $\chi^2 = 26.437$, $df = 16$, $p = 0.0408$, RMSEA = 0.056, CFI = 0.956, AIC= 4093.519. The single change to the significance pattern compared to full results was that there is significant $c_{11females}$ for number of lifetime sexual partners, which was not significant when DZ-OS twins were included.

Table 3.8

Standardized path loading for bivariate models excluding OS twin pairs

	a11	a21	a22	c11	c21	c22	e11	e21	e22
Female	0.304	0.731	0.008	0.521	0.283	0.006	0.798*	0.333*	0.523*
Male	0.439	0.676	0.296	0.378	0.328	0.001	0.815*	0.314*	0.498*
Constrained	**0.292**	**0.651**	**0.000**	**0.496**	**0.344**	**0.000**	**0.818***	**0.359***	**0.573***

Note. **Bold** indicates the best fitting model, fit statistics: $\chi^2 = 66.026$, df = 53, p=0.1079, RMSEA = 0.0034, CFI = 0.972. Path estimates are similar in magnitude to full results, though four previous paths ($a_{11males}$, $a_{21males}$, a_{21}, & c_{11}) fell from significance when excluding DZ-OS pairs.

Table 3.9

Standardized path loadings for trivariate modelsexcluding OS twin pairs

		11	21	31	22	23	33
Female	**a**	**0.746***	**0.441***	**0.632***	**0.021**	**0.343**	**0.097**
	c	**0.299**	**0.017**	**0.159**	**0.435***	**0.244**	**0.001**
	e	**0.595***	**0.130**	**0.198***	**0.774***	**0.296***	**0.511***
Male	**a**	**0.607***	**0.128**	**0.307**	**0.416**	**0.580**	**0.292**
	c	**0.626***	**0.381**	**0.362***	**0.014**	**0.008**	**0.002**
	e	**0.491***	**0.242***	**0.157***	**0.779***	**0.283***	**0.499***
Constrained	a	0.806*	0.069	0.321*	0.214	0.470*	0.350
	c	0.290*	0.583*	0.411*	0.015	0.007	0.001
	e	0.516*	0.490*	0.350*	0.608*	0.166*	0.484*

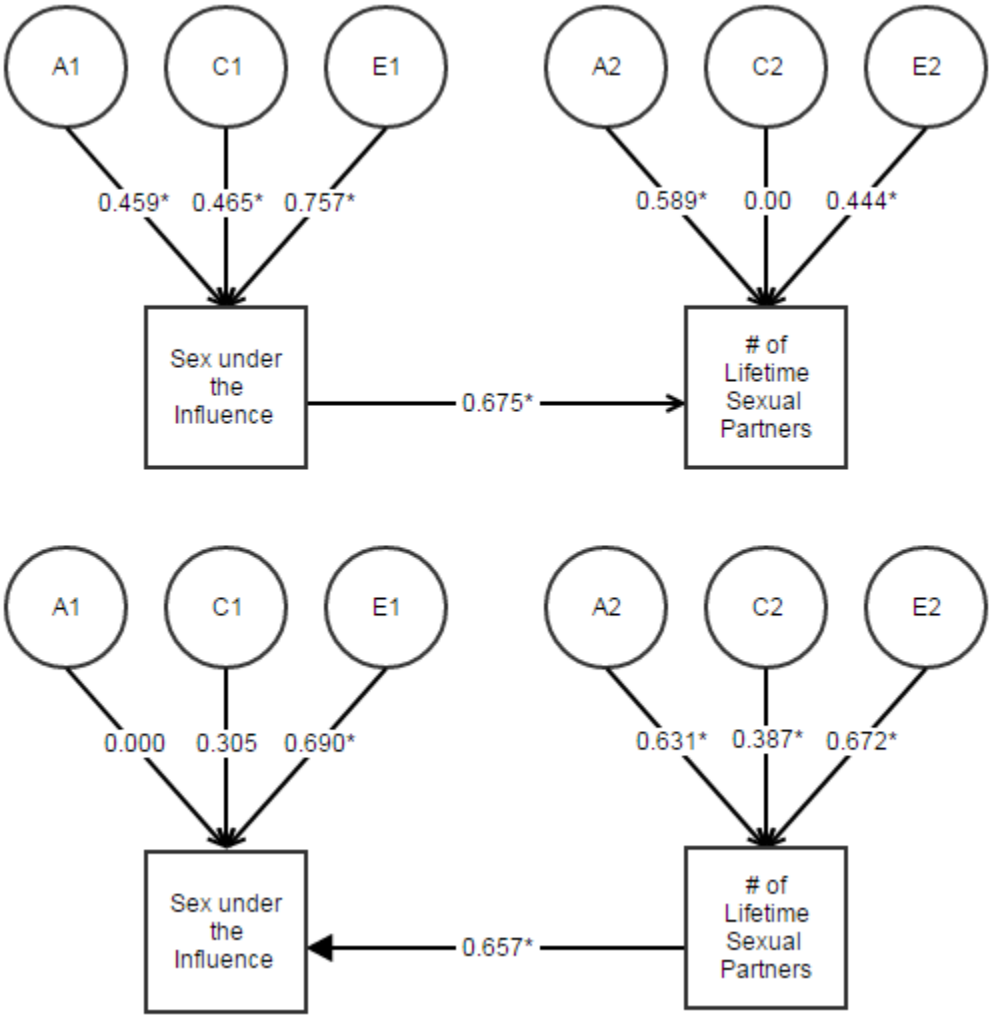
Note. **Bold** indicates the best fitting model, fit statistics: $\chi^2 = 128.247$, $df = 111$, $p = 0.1257$, RMSEA = 0.027, CFI = 0.989. Path estimates are similar in magnitude to full results, though four previous paths ($c_{11females}$, $c_{31females}$, $c_{21males}$, $e_{21males}$, $e_{31males}$, $c_{22males}$, $c_{23males}$, a_{21}, a_{31}, a_{23}) fell from significance when excluding DZ-OS pairs. Four non-significant became significant when excluding OS pairs ($c_{22females}$, $e_{23females}$, $e_{22males}$, c_{21}).

Direction of Causation Models

The direction of causation models are nested under the best fitting bivariate twin model (e.g. separate thresholds for sex under the influence, constrained means for number of lifetime partners, and constrained path estimates across males and females; See Figure 3.4). In the model testing whether sex under the influence has a direct effect on lifetime number of partners, constraining A, C, and E covariance through a single causal pathway resulted in a significant decrement of fit of the model ($\chi^2_{diff}(2)= 22.629$, $p<.001$). This suggests this direction of causation is unlikely, and the observed patterns of covariance are unlikely to be due to mediated pleiotropy or mediated shared environmental influences. Interestingly, the model testing reverse causation (e.g., whether an increase in lifetime number of partners increased the likelihood of sex under the influence) did not result in a significant decrement of fit ($\chi^2_{diff}(2)= 2.995$, $p=.2237$).

Figure. 3.3

Direction of causation models with standardized path loadings constrained across gender



Discussion

In this study, we replicated the association between drug and alcohol use and risky sexual behavior (RSB) using a population-based, genetically informative sample. Higher scores on our composite measure of sex under the influence (e.g. how frequently substance use influenced sexual decision making) predicted higher numbers of lifetime sexual partners. Additionally, we found evidence that people who use more drugs and alcohol generally, also have higher numbers of partners. Multivariate twin analyses were used to model the shared genetic, common

environmental, and nonshared environmental influences shared between sex under the influence and RSB (i.e., common liabilities). As both additive genetic and nonshared environmental influences were significant sources of covariance between sex under the influence and number of lifetime partners; mediated pleiotropy could not be ruled out and causality could not be inferred. Thus, we tested the likelihood of causation using both bivariate twin models and direction of causation models.

Our models were able to decompose sources of variance and covariance into additive genetic, shared environmental, and nonshared environmental factors. If there were no familial influences on sex under the influence there would be overlapping genetic and shared environmental influences on these traits (e.g. rA and rC would be equal to 0) and biological pleiotropy direct effects of shared environment could be ruled out. Results from the bivariate model are inconsistent with this scenario. Evidence for additive genetic covariance (and to a lesser extent marginal evidence for shared environmental covariance) was found in the combined model, where female and male parameters were set to be equal. As both additive genetic and nonshared environmental influences were significant sources of covariance between sex under the influence and number of lifetime partners; mediated pleiotropy could not be ruled out and causality could not be inferred. Additionally, we cannot be certain how much of the non-shared environmental covariance is due to correlated liabilities (unique environmental influences on both behaviors not shared by twins) or spurious correlated measurement error, compared to how much of rE is measuring a true casual effect. Thus, the bivariate model was inconclusive about causality.

We hypothesized that the some of the overlap between sex under the influence and RSB may be related to an overall propensity towards drug use generally, so we examined general

substance use as a potential mediator. A stringent test of non-causation between sex under the influence and number of lifetime sexual partners would have been supported if the non-shared environmental overlap between these traits reduces in significance once controlling for general substance use. In this case, it can be assumed that these nonshared environmental influences (i.e. environments that increase the likelihood for drug use both within and outside the context of sexual activity and RSB itself), are reflective of third variable confounding rather than causation. Evidence of shared environmental influences on all three behaviors reduces the likelihood that these influences are only acting on lifetime number of partners through sex under the influence. This pattern was identified in females only, though there are some noteworthy caveats. In the trivariate model the covariance was explained by the first common E factor loading on all three traits, though the path estimate loading on sex under the influence was not significant. This pattern of non-significant overlap may be due to reduced power, as the use of sex limitation models reduces the number of data points contributing to a given path estimate (i.e., the trivariate analyses reduce the effective sample size from [n=1870] to females [n=1047] and males [n=823]). Although all subjects are included in unconstrained and constrained models, the information to estimate sex-specific parameters in a model is limited to data for a given sex. It was also necessary to include two ordinal measures in our study; thus, our models have significantly reduced power compared those using continuous variables (Neale, Eaves, & Kendler, 1994).

The patterns of results from the Cholesky decomposition were inconclusive, and additional direction of causation (DoC) models were required for causal inference. These models tested the likelihood that: 1) sex under the influence has a direct effect on lifetime number of partners, and 2) reverse causation, or that RSB actually directly causes more sex under the

influence. We were unable to test a third theoretical model, reciprocal causation, as our RSB measure did not have multiple indicators. Results suggested that sex under the influence did not have a direct causal effect on lifetime number of sexual partners, rather the reverse pattern better fit the data.

As the aim of the study was to test whether drug and alcohol use during sexual decision-making caused people to take more risks, the reverse causation finding is somewhat unexpected. The underlying mechanisms of this potential causal pathway were not explored in this study and are avenues for future research. For instance, it is unclear whether these effects are driven by sensation seeking (e.g. those with higher partners may be more motivated to add drugs or alcohol into the situation to enhance the experience) or partner effects (e.g., those with higher RSB are more likely to acquire partners who are more likely to introduce substance use into the sexual situation. Unmeasured third variables cannot be completely ruled out and should be investigated further.

Previous studies that used biometrical models to test the role of adolescent alcohol and drug use on other RSB cast similar doubts on a causal model in adolescence. One study found that age of smoking and drinking were non-causal predictors of early age of sexual initiation (e.g. shared vulnerabilities for both traits existed, controlling for conduct disorder symptoms), though tests of early drunkenness could fit a *partially causal* model (Deutsch et al., 2014). Agrawal et al. used a sample of female twins to test a causal model of adolescent cannabis use (e.g. before age 17) on repeated voluntary unprotected sex in early adulthood (2016). After controlling for related behaviors and covariates including adolescent drinking and cigarette smoking, early sexual initiation, early puberty, symptoms of conduct disorder, and childhood sexual abuse, some genetic overlap was detected (i.e., genes that influenced early cannabis

initiation also influenced adult non-condom use via biological or mediated pleiotropy). As genes did not explain all of the covariance between these behaviors, the authors did not rule out the possibility of a *partial causal* role between these behaviors (Agrawal et al., 2016). To our knowledge, there are no genetically informed studies that test a measure of sex under the influence (substance use during sexual decision making) during adolescence. Given that this is a target population for intervention, a logical follow up would be to extend this specific test to younger samples.

There are several ways in which our study differs from past research, which should be considered when interpreting the results. Our sample was community based, therefore the range of behavior was normal and extreme risky behavior was rare. Caution should be taken when extending our results to clinical populations or those with higher mean levels of sexual partners. Second, we sampled our participants in early adulthood when approximately half of the sample was in a committed or monogamous relationship. In contrast, many of the other investigations have focused on late adolescence. While late adolescence is often a period of radical environmental changes (e.g. leaving parent's home), early adulthood is particularly interesting in terms of selecting different life trajectories with various degrees of risk (e.g., getting married). Additionally, our use of the sex under the influence composite differs from some of the measures used in other studies. In essence, it is an attempt to measure how much drug and alcohol use influences sexual decision-making, rather than a measure which objectively assesses whether drug and alcohol led to risker sex than would have occurred if sober. Possible implications of this measure (e.g., subjectivity, contributions to gender differences) are discussed below. Finally, it is possible that a minority of the participants with only one partner in the past 5 years scored as "no risk" would have endorse some of the sex under the influence items (e.g. had a drink to feel

more comfortable with [regular] partner). Consistent with other work (Epstein et al. 2014), we chose to consider these behaviors within the context of a committed relationship as non-risky due to several reasons. While sex under the influence may be risker with a regular partner in some regards (e.g. unintended or unwanted pregnancy), on average the risk for contracting or spreading an STI from a regular partner is minimal compared to people with multiple partners. This risk is highest for those with concurrent multiple partners (Morris, & Kretzschmar, 1997); however, risk is still elevated for serial monogamists or those with multiple subsequent partners (Corbin & Fromme, 2002). Stated again, the sex under the influence measure aimed to measure the frequency of situations where drug and alcohol use potentially led to risker sexual behavior, which is markedly different than when a person in a committed relationship may use drugs or alcohol to feel more relaxed or comfortable with their regular partner.

Despite these caveats, the gender differences in all of the models are of particular interest. It is clear that the reasons for drinking or using drugs and alcohol before (or during) sexual activity are different across males and females. In males, sex under the influence was more highly heritable and there is substantial genetic overlap specific to sex under the influence and sexual risk behavior. As such, future research should further explore the biological systems core to these behaviors (e.g. hormonal system, genetic basis of underlying traits such as sensation seeking). In females, genes may influence both substance use within and outside of sexual activity and increased sexual encounters. While the alcohol myopia hypothesis (Steele and Josephs, 1990) posits that the link between substance use and RSB is due to direct physiological impairment in information processing and decision making, differential social pressures for men and women to use substances or limit sexual behavior may also contribute to these gender differences. A self-fulfilling prophecy effect (Lang, 1985) can occur when people have strong

social expectations about drug and alcohol use. For example, Dermen and Cooper found that among those with strong expectations that drinking alcohol promotes sexual risk taking, alcohol use was negatively correlated with condom use during first intercourse and first intercourse with most recent partner (2000). Existing gender differences in social expectations or motivations for alcohol and drug use could explain these results (Beck, Thombs, Mahoney, & Fingar, 1985). Additionally, there may also be different pressures for males and females to *report* that their behavior was a consequence of alcohol or drug use. For instance, to the extent to which there is pressure men to have more partners and women fewer partners and there are strong social expectations about substance use and risky sexual behavior (e.g. in college campuses; there may be some benefit for women, especially, to endorse that their decisions were clouded by drug or alcohol use), we may expect different patterns of sex under the influence and RSB for males and women. As such, our measure may be capturing the frequency in which substance use impaired sexual decision making (presumably, leading to increased risk), capturing a tendency to use substances in order to facilitate sexual encounters, or a capturing a tendency to endorse sex under the influence measures as a form of post-hoc rationalization to remove dissonance or social shame of RSB. Further investigation into the specific motivations to endorse individual risky behaviors associated with sex under the influence (e.g. non-condom use while intoxicated) could inform our interpretation of these gender differences.

In sum, findings indicate that claims that drug and alcohol use during sex causes risker sexual behaviors may be overstated. We identified no clear pattern of causality using bivariate twin modeling, and found evidence that a large proportion of the shared liabilities between sex under the influence and number of lifetime sexual partners was substance use more generally. DoC models indicated that not what it unlikely that sex under the influence caused more RSB,

but that the reverse may be true. Future research is required to investigate the nature of this potential causal pathway.

CHAPTER IV

ISSUES IN MEASURING SEXUAL BEHAVIOR AND DEFINING RISKY SEXUAL BEHAVIOR

Background

This chapter discusses the issues relevant for measuring sexual behavior and defining risky sexual behavior (RSB). These considerations are important for the following reasons, 1) to inform and clarify the meaning and interpretation of the variables selected for analysis in Chapter III (i.e. number of lifetime sexual partners and sex under the influence), 2) to assess the utility of our primary measure of RSB (i.e., M-RBQ) and to create a revised version, and 3) to select an optimal phenotype to be used in later genetic analyses (i.e., in Chapter V).

Measuring Sexual Behavior

Not long after the seminal Kinsey Reports of human sexuality were published (Kinsey, Pomeroy, & Martin, 1948; Kinsey et al., 1953), numerous academics expressed concern regarding sampling and measurement issues relevant to sexual behavior. Such critiques are briefly mentioned here, as well as a discussion of newer approaches that may increase the accuracy of measurement of sexual behavior.

Social Desirability

Of primary concern is social desirable responding (Bernreuter, 1933), or the tendency for participants to report a modified response for the purpose of "looking good." Social desirability effects may be especially relevant for items of sensitive nature. Factors that influence whether or not a participant self-discloses information that may be emotionally distressing or whether self-disclosure provides an accurate representation of behavior may depend the specific questions

asked, the context of the interview, and the individual participant's characteristics (Catania, 1999).

Item Selection

The specific questions asked on an assessment have the potential to bias reporting. Reporting will be influenced by the language (e.g., does the question provoke an emotional response?), topic sensitivity (e.g., are stigmatized behaviors or violation of societal norms assessed?), and question clarity (e.g., does the interviewee understand the question?) (Catania 1999). These biases may be influenced by characteristics of the participants, including participant age, gender, or race (Fenton et al., 2001) For instance, due to widespread disagreement about what constitutes the act of "having sex," even presumably straight forward items can suffer from sample bias if clear operational definitions are not provided (Young, Palacios, & Penhallow, 2012). For example, men may be more likely to include non-penetrative sex partners compared to women (Catania et al., 1996).

Recall

Another major source of measurement error can be caused by recall bias. Events that are distal (Croyle & Loftus, 1993) or are frequent in nature (Catania, Gibson, Chitwood, & Coates, 1990) are more likely misreported than those that are recent and rare. Recall issues may also vary across different types of participants. For instance, white females and participants with more sexual partners were more consistent in their recall or reporting of age of sexual initiation across multiple waves of data collection (Goldberg, Haydon, Herring, & Halpern, 2014).

To reduce bias, several methods have been suggested. For tracking sexual behavior in a short period of time, use of daily diary tracking may be used to record sexual behavior as close to the event as possible (Graham et al., 2003). In longitudinal studies, aids can be used to prompt

memory recall about a specific time period or event. For example, calendar interviewing approaches can improve recall by prompting participants to record significant life events or contextual personal histories as scaffolding for the event being recalled (Freedman et al., 1988).

Computer-Assisted Self- Interviews

Given the sensitive nature of items assessed on surveys of RSB, it has long been a standard to collect this information with private, self-administered surveys. The use of computer-assisted self-interview can increase efficiency of data collection (Turner et al., 1998; Mustanski 2001; Mercer 2010) and improve overall accuracy (Schroder, Carey & Vanable, 2003b). For example, computer-assisted self-interviews reduce issues with social desirability and disclosure, while the ability to program skip-out and routing patterns reduces the need for complex instructions (Fenton et al., 2001).

Several experiments have explored the variables that may influence reporting of risk behavior or self-disclosure, and use of computer assisted self-interviews provides an easy way to conduct such studies (Mustanski 2001). Results of these studies suggest potential gender differences in self disclosure by survey type. While there was no difference in reporting for women, men reported higher sexual risk taking when assessed via a structured interview than compared to a self-administered survey (Turchik & Garske, 2009).

Generalizability

Studies of sexual behavior have typically utilized the following study designs or ascertainment strategies: population samples, representative samples, selected samples (e.g. high risk groups), partner or network studies, and ethnographic or qualitative studies (Fenton et al., 2001). Caution should be taken when inferring across sample group types. For instance, high risk, low prevalence behavior may not be well captured in a general population sample.

Similarly, inclusion of high risk participants may identify predictors, correlates, and health outcomes that are only specific to smaller subgroups of the populations (e.g. high risk groups such as injection drug users, men who have sex with men, etc.). Access to high risk or understudied populations is increasingly feasible due to the rise in computer assisted self-interviews for sexual behavior inventories and internet based surveys (Mustanski 2001), as well as access to larger overall sample sizes (Mercer 2010). However, given the strong possibility of floor and ceiling effects of particular behaviors, variability within any particular sample should be carefully considered (Leigh & Stall, 1993).

Existing questionnaires and surveys

There is no consensus on a "gold standard" for assessing sexual behavior, though many measures have been proposed (Fisher, Davis, Yarber, & Davis, 2013; Mirzaei, Ahmadi, Saadat, & Ramezani, 2016). Some large scale studies, such as the ongoing Youth Risk Behavior Surveillance System (YRBSS) conducted by the Center for Disease Control (Brener et al., 2013) often use short scales measuring a few established predictors of unintended pregnancies and sexually transmitted infections. Other more detailed scales have been developed to target specific age groups or more nuanced behaviors. Several of these measures are discussed below.

The Sexual Risk Behavior Beliefs and Self-efficacy Scales (SRBBS) were created to assess behavior in a school-based program for HIV. The SRBBS includes scales to measure attitudes, norms, self-efficacy and barriers to condom use (Basen- Engquist et al., 1998). These scales have been found to be reliable with good concurrent validity for high school students (Basen- Engquist et al., 1999).

The Sexual Risk Survey (SRS) was created to be a valid, comprehensive measure of sexual risk in college populations (Turchik & Garske, 2009). Principal component analysis

identified five distinct factors of sexual risk including 1) sexual risk taking with uncommitted partners, 2) risky sexual acts (i.e., unprotected sex or sex under the influence), 3) impulsive sexual acts, 4) intent to engage in risky sexual behaviors, and 5) risky anal sexual acts.

As previously mentioned in Chapter II, we administered the Modified Risk Behavior Questionnaire (M-RBQ, adapted from Booth, Corsi & Mikulich-Gilbertson, 2004). This measure was designed to assess past year risky sexual behavior in adulthood; however, the measure suffers from several of the measurement issues discussed here. A full description of the limitations of the M-RBQ are listed below, as well as a description of changes made for a new version of the questionnaire.

Defining RSB

We limit the definition of risky sexual behavior (RSB) to those behaviors that increase the risk for contracting human immunodeficiency virus (HIV), other sexually transmitted infections and diseases (STI/STDs), or unwanted pregnancy. By limiting the definition of risky behavior to those leading to negative health outcomes (McKie et al., 1993), researcher bias about which behaviors are deviant, abnormal, or moral is reduced. Other potential maladaptive outcomes related to sexual behavior (e.g. increased risk for psychopathology or other distress, financial hardships, etc.) have sometimes been suggested in the literature, though many of these associations suffer from implicit untested causal assumptions. Finally other established studies such as the YRBSS (Brener et al., 2013) use a similar limited definition focusing on health-risk behavior.

In 2015 alone, the United States had approximately 20 million new STI/ STDs cases (CDC, 2016). Beyond the direct effects of acquiring a STI/STD, indirect health consequences include an increased risk for complications during pregnancy or conception (e.g. ectopic

pregnancy, premature birth of child, pelvic inflammatory disease, etc.), infertility, certain cancers (e.g. cervical, penile, Kaposi's sarcoma, etc.), and neurological syndromes (Aral, 2001).

While rates of unwanted pregnancies can be difficult to measure, the National Survey of Family Growth (NSFG) has estimated approximately 49% (i.e., 3.2 million out of 6.7 million) pregnancies reported in the United States in 2006 were *unintended* (Finer & Zolna, 2011). In comparison to planned pregnancies, unintended pregnancy increases the risk for illness and mortality for both the mother and fetus (Gipson, Koenig, & Hindin, 2008). Rates of unintended pregnancy are correlated with age, with highest rates occurring in teenage pregnancies (Finer & Zolna, 2014). Though rates of teenage pregnancy (occurring in females ages 15-19) has been declining in recent years in the United States, recent rates remain much higher than those of other developing countries (AGI, 2016) and the proportion of those pregnancies that are unintended remained high (Zolna & Lindberg, 2012). Abortion or adoption can be used as a rough proxy to determine how many unintended pregnancies were indeed unwanted; however, access to these services can vary widely by age, race, socio-economic status, or geographical location (Henshaw, 1995; Meier et al., 1996).

Given this definition of RSB, the most commonly measured behaviors in the literature include number of sexual partners, age of sexual initiation, frequency of condom use, sex under the influence of alcohol or drugs, risky sexual acts (e.g., those with higher rates of transmission of diseases or infections), or sex with risky partners (e.g. with higher risk of disease or infection, possibly casual partnerships).

Number of partners

Some measures of number of partners include only recent partners (e.g., partners in past year) or number of partners during riskiest period of sexual behavior, but the most commonly

used measure is a number of lifetime sexual partners. While there is not a consensus threshold that indicates risk, having a higher number of sexual partners is a strong predictor of STI/STDs (Karlsson, et al., 1995; Santelli et al., 1998; Sturdevant et al., 2001; Epstein et al., 2013). Though having zero sexual lifetime sexual partners may be related to negative social and interpersonal consequences (Gesselman, Webster & Garcia, 2017), the low end of this spectrum is not considered risky in that it does not confer risk for STI/STDs and unwanted pregnancy.

Mean number of lifetime sexual partners for any particular population is influenced by several factors. Strong generational effects influence this average as norms about marriage, sex for procreation, and attitudes about HIV/AIDs shift across time (Twenge, Sherman, & Wells, 2015). While cohort is an important consideration, age will also necessarily be related this variable as counts can only increase or remain the same over a lifespan. Gender differences in self-reports are common (Morris, 1993). Male over reporting and female under reporting may be due to gender-related social expectations or differential estimation strategies (Brown & Sinclair, 1999). As a final note, median statistics or ordered categories are commonly reported given positive skew in the distribution. Outliers can more easily bias the mean. Median number of lifetime sexual partners as measured by the 2006-2008 National Survey of Family Growth were 1.4, 2.6, and 3.6 for females aged 15-19, 20-24, and 25-44, respectively, and 1.8, 4.1, an 6.1 for males aged 15-19, 20-24, and 25-44 (Chandra et al., 2011).

Age of initiation

Mean age of sexual initiation is typically around 17 years of age (CDC, 2013), with small effects of gender, type of sex act, or sex of partner (CDC, 2013). *Early* sexual initiation is associated with an increased risk for STI/STDs, higher number of sexual partners, higher number of *risky* sexual partners, and sex under the influence (Sandfort, Orr, Hirsch, & Santelli, 2008).

Definitions of sexual initiation often include any initiation of anal, oral, or vaginal sex, though individual studies vary the criteria for what is considered to be early sexual initiation (Heywood, Patrick, Smith, & Pitts, 2015). Early initiation has been defined as from before the age of 15 (Meier, 2007; Springs & Halpern, 2008; Epstein et al., 2013), before the age of 16 (Paul, Fitzjohn, Herbison, & Dickson, 2000; Cavazos-Rehg et al., 2008; Huibregtse et al., 2011), or relative to the sample mean or distribution (Bricker et al., 2006; Sandfort, Orr, Hirsch, & Santelli, 2008). Finally, age of initiation is one of the most widely available measures across studies (Zimmer-Gembeck & Helfand, 2008).

Notably, there is also evidence to suggest that late age of sexual initiation may also be associated with a number of health outcomes such as sexual problems (arousal in both sexes, erectile or orgasm problems in men; Sandfort, Orr, Hirsch, & Santelli, 2008). Some distinctions should be drawn between those who choose to abstain from or delay sexual activity in line with personal beliefs to those who unable to initiate (Boislard, Bongardt, & Blais, 2016), whereas the predictors of late initiation are different in self-selectors.

Frequency of condom use

Measuring condom use is complex, in that condomless sex is not necessarily always risky but is often used as a proxy for exposure to STI/STDs or potential for unintended pregnancies. Measures may range from general condom use (e.g., yes or no), frequency of condom use, to condom use during a specific encounter (e.g., at sexual initiation or at last sex). However, there are a number of contextual factors that may moderate risk such as partner type (e.g. causal vs. regular partner), specific sex act (e.g. anal, oral, or vaginal), and the frequency of sex. Further considerations about whether or not a heterosexual individual is actively trying to get pregnant or is using an alternative form of birth control are often ignored by these measures. A meta-analysis

of 56 studies of self-reported condom use suggests that addressing these issues has improved over the past few decades; however, a number of measurement issues still are widely pervasive (Noar, Cole, & Carlyle, 2006) and correlations with other RSB measures can depend on the type of variable used (e.g., count vs. relative frequency measure; Schroder, Carey, & Vanable, 2003a).

As such, the pattern of correlations between condom use measures and other RSB measures is not always consistent. Several studies find that inconsistent condom use or condomless sex is correlated with having multiple and or risky partners (Catania et al., 1992), sex under the influence (Oshri et al., 2013), and a general measure of impulsiveness (Eysenck Impulsivity Scale; Kahn, Kaplowitz, Goodman, & Emans, 2002). However, there is also conflicting evidence suggesting that condomless sex is not always correlated with other risky behaviors. Results from a large-scale study in Britain (n=13,765) suggests that people are *more* likely to report condom use at last sex when circumstances were more risky (e.g., sex with a causal partner, when having high numbers of recent partners, etc; Cassell, et al., 2006). Another study of adolescent risk behavior found inconsistent condom use occurred in both adolescents with low and high number of sexual partners according to a longitudinal latent profile analysis (Beadnell et al., 2005). Thus, condom use may not predict other types of RSB or predict safer sexual behaviors in some contexts. Condom use as a proxy of risk behavior should not be considered in isolation and may be a less direct measure of RSB.

Sex under the influence

Estimates from the National Epidemiological Survey on Alcohol Related Conditions (NESARC) suggest that over four million American adults regularly consume alcohol before sex (Eaton et al., 2015). Use of drugs or alcohol during sex has long been considered a potential

factor that may increase risk taking (Leigh & Stall, 1993). While it is a consistent correlate of other forms of RSB, meta-analyses exploring whether or not sex under the influence causes an increase in STI/STD or unwanted pregnancy have had mixed results (Rehm, Shield, Joharchi, & Shuper, 2012; Shuper, Neuman, Kanteres, Baliunas, Joharchi, & Rhem, 2010; Morrison, Gillmore, Hoppe, Gaylord, Leigh, & Rainey, 2008; & Leigh, Vanslyske, Hoppe, Rainey, Morrison, & Gillmore, 2008).

Risky sex acts

Differential sex acts confer differential risk for disease. For example, risk for acquiring HIV is highest during receptive anal sex, and decreases in likelihood across the following acts: insertive anal sex, receptive vaginal sex, insertive vaginal sex, and finally oral sex (Patel et al., 2014). This has led some researchers to make distinctions between specific types of sexual acts as more or less risky. As risk of exposure increases as frequency of unprotected sexual activity increases, frequency of specific risky sexual acts could magnify the risk of inconstant condom use.

However, the long-term consequences of certain sex acts may be disproportionate for certain populations. For example, women experience a disproportionate amount of the health consequences of STI/STDs (CDC, 2011). Also, men who have sex with men who are more likely to engage in more risky sexual acts (in terms of rate of transmission) and potentially with riskier sexual partners (Valleroy et al., 2000; Koblin et al., 2003).

Risky sexual partners

Risky sexual partners are those with STIs/STDs or HIV, or those with high risk for these diseases. Proxy measures for risky partners include injection drug users, those participating in risky sexual behavior listed above, or other characteristics of the partner that increase the risk for

negative health outcomes. For example, characteristics of the partner or the partner type may be predictive of risk (e.g. partner age difference or context of relationship). Indeed, females who initiated with *older* sexual partners compared to those with same aged partners have worse STI/STD outcomes and higher risk for unwanted pregnancy (Ryan, Franzetta, Manlove, & Schelar, 2008). Measures of partner type in the literature often make a distinction between regular (e.g. ongoing, romantic, or committed partnerships) and causal (e.g., non-romantic, or uncommitted partnerships). While there has been some work linking certain partner types to well-being and satisfaction outcomes (Harden & Mendle, 2008; Higgins et al., 2011), the link between partner type and negative health outcomes are less clear.

There is often an assumption that casual partnerships are risker, given less available information about disease status or sexual history. Even when such matters are discussed, there is a second assumption that a casual partner may be less honest than a regular, trusted partner. Indeed, a significant proportion of people admit they would lie about disease status or sexual history when asked by a potential sex partner (Cochran & Mays, 1990). However, such assumptions can lead to differential patterns of condom use by partner type (Macaluso, Demand, Artz, & Hook, 2000; Lescano et al., 2006) where condoms are more likely to be used with casual partners and less likely to be used with regular partners. This pattern can create a potential vulnerability and risk for disease exposure, especially in the case of serial monogamy (i.e. having many subsequent "regular" partners (Corbin & Fromme, 2002).

Some studies have considered factors such as monogamy, polyamory (i.e. consensual non-monogomy or open relationships), extra-marital relationships, or frequency of cheating to predict negative health outcomes. Indeed, STI/STD risk is highest for those with concurrent multiple partners (Morris, & Kretzschmar, 1997). It is possible that concurrent partnerships

increase negative health outcomes only because they necessarily increase number of sexual partners (Lurie & Rosenthal, 2010), or they increase negative outcomes because they make it more likely that those partnerships are with others who have multiple concurrent partners. However, contextual factors about these partnerships may mitigate potential risks. For example, there is some evidence that individuals with secondary partners (e.g. likely, extramarital partners) are *more* likely to use condoms outside their relationships compared to with their regular partners (Choi, Catania, & Dolcini, 1994). Similarly, 99% of 343 participants in a polyamory study had formal agreements about disclosure and relationship rules with partners, which may help reduce risk of negative health outcomes (Wosick-Correa, 2010). In contrast, undergraduate students with multiple serial partnerships were less likely to disclose information about previous partnerships and previous inconsistent condom use, but were no less likely to disclose STI/STD status than those with only one recent partner (Desiderato & Crawford, 1995). However, it is unclear how people actually modify their behavior, such as condom use, in different relationship contexts.

Selecting an Optimal RSB Phenotype

These efforts to define and measure RSB have a larger goal: to identify an optimal phenotype for genome wide analyses conducted in Chapter V. Our selection criteria are as follows: 1) moderate to high heritability, 2) reliable with low measurement error [i.e. easily standardized across samples], 3) little heterogeneity or sex-limitation, and 4) a continuous and normal distribution (Freimer & Sabatti, 2003; Manchia et al., 2013). Finally, the phenotype should be predictive negative health outcomes, in accordance with our definition of RSB.

Methods

Sample

Multiple samples were used to explore measurement issues in RSB (i.e., genetically informed samples and a pilot focused on refining our phenotypes).

The larger sample was drawn from the CADD samples (previously described in Chapter II). These samples include several family based designs including twin studies (i.e., Colorado Longitudinal Twin Study [n=922] and the Colorado Community Twin Study [n= 2013]), the Colorado Adoption Project (n=195), and a family study of "clinical" probands in treatment for adolescent antisocial drug dependence, their siblings, and matched control families (n=750); Stallings et al., 2003). Thus, the aggregate sample is enriched, or includes participants selected for behavioral disinhibition, as elevated conduct problems and substance dependence symptoms were used as inclusion criteria for "clinical" probands (for additional criteria and description of recruitment, see Stallings et al., 2005). The final sample (n=3380, females= 49.9%) was drawn from the most recent waves of data collection in early adulthood (*m*= 26.25 years, *s.d.*= 3.39) to capture lifetime risky sexual behavior.

For the pilot portion of the study, participants were recruited were undergraduates recruited from the Department of Psychology subject pool at the University of Colorado Boulder (n= 182, females =64.29%). Participants under age 18 were excluded from the survey, resulting in an age range from 18-25 (*m*=18.73 years, *s.d.*= 1.22). Participants completed the study in partial fulfillment of the research requirement for an introductory psychology course.

Measures

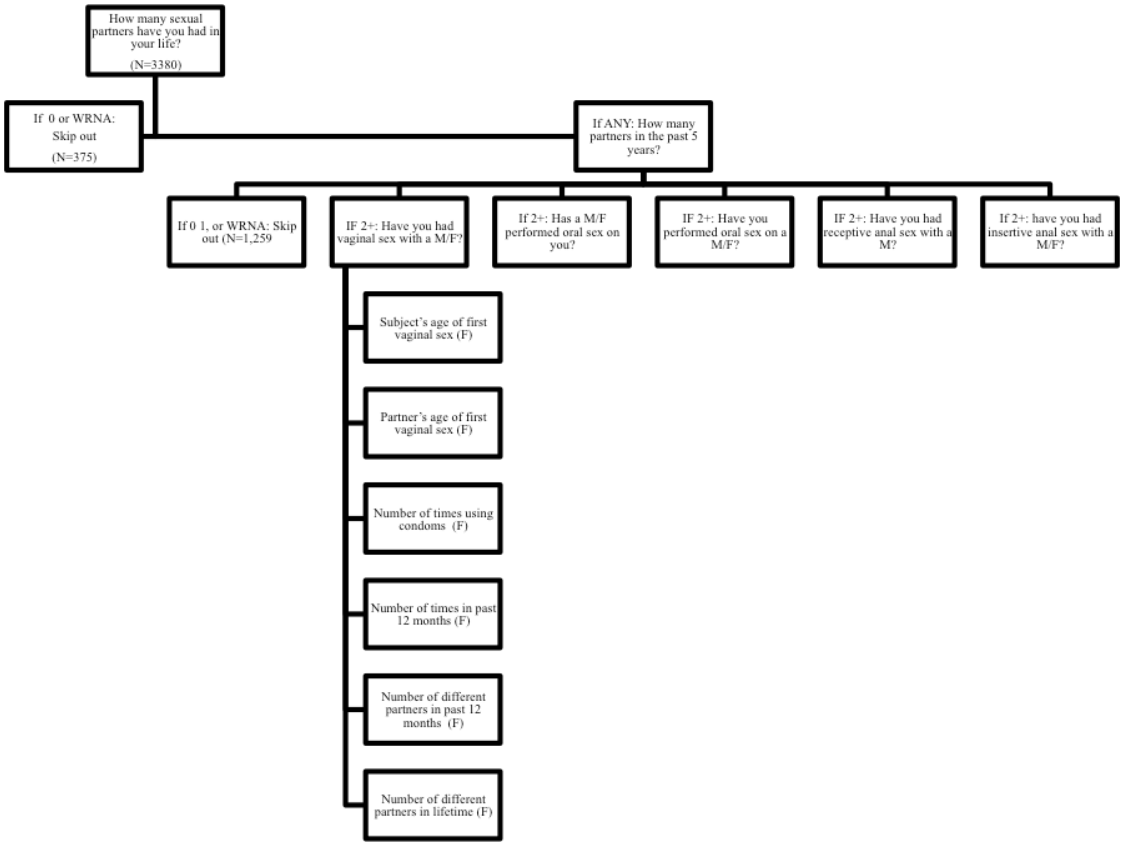
RSB was assessed using the modified Risky (M-RBQ, adapted from Booth, Corsi, & Mikulich-Gilbertson, 2004). Items included the number of sexual partners in the past five years, lifetime number of sexual partners, as well as a series of questions based on endorsement of the following specific types of sexual activity depicted in Figure 4.1. For each specific sex act

endorsed with a male or female partner, participants were asked age of initiation, their partner's age of at initiation, number of different (act specific) partners in their lifetime, number of different (act specific) partners in past 12 months, frequency of engaging in each specific sex act in past 12 months, and frequency of using condoms or other protection during specific sex act in past 12 months). Sex acts included anal, oral, and vaginal sex.

A three-tiered skip-out pattern was applied to reduce the burden on participants; and required by the University of Colorado IRB. First, non-sexually active participants (i.e., participants who reported zero sexual partners) were not administered the M-RBQ. Second, sexually active participants reporting zero or one partner in the past 5 years also skipped out of further questioning because they could be considered low risk. This tailored -assessment was intended to assess only participants who were engaging in sufficient sexual activity to meaningfully assess RSB. Finally, participants reported on the gender of their sex partners. Subjects only reporting male partners, for example, were not asked questions about sex with females. Only subjects reporting both male and female sex partners were asked all items. This alleviated the need to ask detailed (and potentially upsetting or embarrassing) questions of subjects who would answer no anyway. That is, given information on the gender of their sex partners, the computerized testing asked participants only questions relevant to their partner responses. For example, if a male reported male sex partners, he would be asked, for example, whether he had ever engaged in specific sex acts (e.g., "have you ever had receptive anal sex with a man?"). This item would not be asked of males reporting only female partners. Participants also skipped out of additional follow up items for any specific sex act they did not endorse (e.g. frequency of act, age of initiation of act, frequency of condom use during act, etc.; See Figure 4.1 for skip out structure). This tailored testing considerably reduced subject burden

Figure 4.1.

Skip out structure and question flow for specific sex acts in the M-RBQ



Note: For simplicity, figure shows branch for only one specific sex act. Similar questions are asked for each specific sex act endorsed, with separate questions for male (M) and female (F) partners. Skip outs include those endorsing would rather not answer (WRNA).

Additionally, items regarding negative health outcomes (e.g., " have you ever been told by a nurse or other health care professional that you had hepatitis?"), sex under the influence (measure discussed further in Chapter III), characteristics about sex partners (e.g., "how many partners were injection drug users?"), risky sexual acts such as history selling sex or trading sex for drugs, participants' perceived risk for HIV, sexual orientation, and sexual satisfaction were collected.

The pilot study used the online platform Qualtrics to conduct confidential survey entitled "College Behavior Survey" (See Appendix B). Core items were selected from the M-RBQ. Additional items regarding drug use from the M-RBQ and general risk behavior items from the Risk Taking Assessment (Nicholson, Soane, Fenton-O'Creevy & Williams, 2005) were added, as well as detailed and open-ended items addressing risk behavior motivations. Extensive skip-outs were built into the survey to minimize irrelevant questions (e.g., participants were only prompted to answer follow-up questions regarding marijuana use if they endorsed using marijuana in the past year). Items with no clear association with risk in the M-RBQ were removed and replaced with clarified questions. An additional item was added to the previous sex under the influence items measuring the frequency of regretting "a sexual encounter because you were drinking or using drugs." Finally, items that addressed motivations for endorsing impaired sex were added. For example, those who endorsed the regret item were further prompted to select whether they did so for the following reasons; 1) "I would not have had sex sober, and I wish I hadn't with that person or at that or at that time", 2) "There were some social consequences related to those encounters (i.e., I am embarrassed, it hurt my relationship with someone or a partner," 3) "There were physical consequences related to those encounters (i.e., unintended pregnancy, new STD)", or 4) an open ended "Other (please explain)" option.

Analyses

Descriptive statistics were primarily completed using *R*, version 3.2.2 (R Core Team, 2015) and using the Statistical Package for Social Sciences (SPSS Inc., 2010). As an exploratory analysis of base rates of potential measures of RSB from the M-RBQ in the CADD sample, response rates were calculated to 1) identify potential sources of measurement error, 2) to identify continuous and normally distributed variables, and 3) to eliminate uninformative

variables (e.g., those with very low prevalence or ambiguous associations with negative health outcomes). Similarly, response rates from the pilot sample were used to infer motivation for endorsing sex under the influence items in the M-RBQ.

Genetic analyses were conducted using the software package Mx (Neale, 1997) to identify a variable with high heritability and little heterogeneity or sex-limitation utilizing the CADD sample. Univariate biometric models were used to estimate heritability and test for effects of gender for the most promising measures of RSB (i.e., met the most criteria for an optimal measure of RSB for genetic analyses).

Results from both samples are integrated into a larger discussion and literature review of measurement issues of RSB.

Results

M-RBQ in CADD sample

The exploratory analysis was used to identify base rates of behaviors and negative health outcomes, missing data patterns, and to understand the extent of genetic influences on any given measure of RSB.

Several factors influenced the sample size of item responses. Due to the sensitive nature of the questions, a "would rather not answer" option was available for each item; this resulted in some missing data. Participants who reported zero lifetime sexual partners were not asked any further questions (n=278, 7.2% of sample). Because the M-RBQ was originally designed to assess recent RSB, sexually active participants reporting zero or one partner in the past 5 years also skipped out of further questioning (n=1,181, 33% of sample). In total, 40.3% of the sample did not complete the full questionnaire. Finally, several items were only asked to a subset of

participants (e.g. condom use frequency during anal sex would only be assessed in participants who endorsed engaging in that specific sex act).

Negative health outcomes

We first explored the base rate of negative health outcomes in our sample (i.e. STI/STD and HIV diagnoses). Participants reported if they have ever been told by a doctor or health care professional that they had the following eight diagnoses: hepatitis B, hepatitis C, gonorrhea, syphilis, genital warts (HPV), chlamydia or nongonococcal urethritis (NGU), tuberculosis, and HIV/AIDs (see Table 4.1. for response rates).

Table 4.1.

Percent of responses for STI/STDs diagnoses

	Never	At least 1	WRNA
Hepatitis B	99.7	0.1	0.2
Hepatitis C	99.5	0.3	0.3
Gonorrhea	98.1	1.7	0.2
Syphilis	99.5	0.4	0.1
Genital Warts (HPV)	93.0	6.7	0.3
Chlamydia or NGU	92.8	6.8	0.3
Tuberculosis	99.4	0.4	0.2
HIV or AIDS	99.9	0.0	0.1

Note: "Would rather not answer" = WRNA

Self-reported base rates for lifetime diagnoses of STI/STDs in the CADD sample were low, compared to estimates of 2013 prevalence estimates of 110 million infections in the United States (Satterwhite et al., 2013). For instance, though other studies have also identified HPV to be the most prevalent infection (Forhan et al., 2009), the lifetime prevalence reported here is only a fraction of other reports (i.e., estimated 26.8% of US women; Dunne et al., 2007). However, discrepancies between self-report diagnoses and behaviors with biomarkers are not uncommon (Zenilman et al., 1995; Van Duynhoven, Nagelkerke, & Van de Laar, 1999; Brown et al., 2012) and biomarkers can be used to reduce errors of socially desirable responding (Gallo et al., 2013).

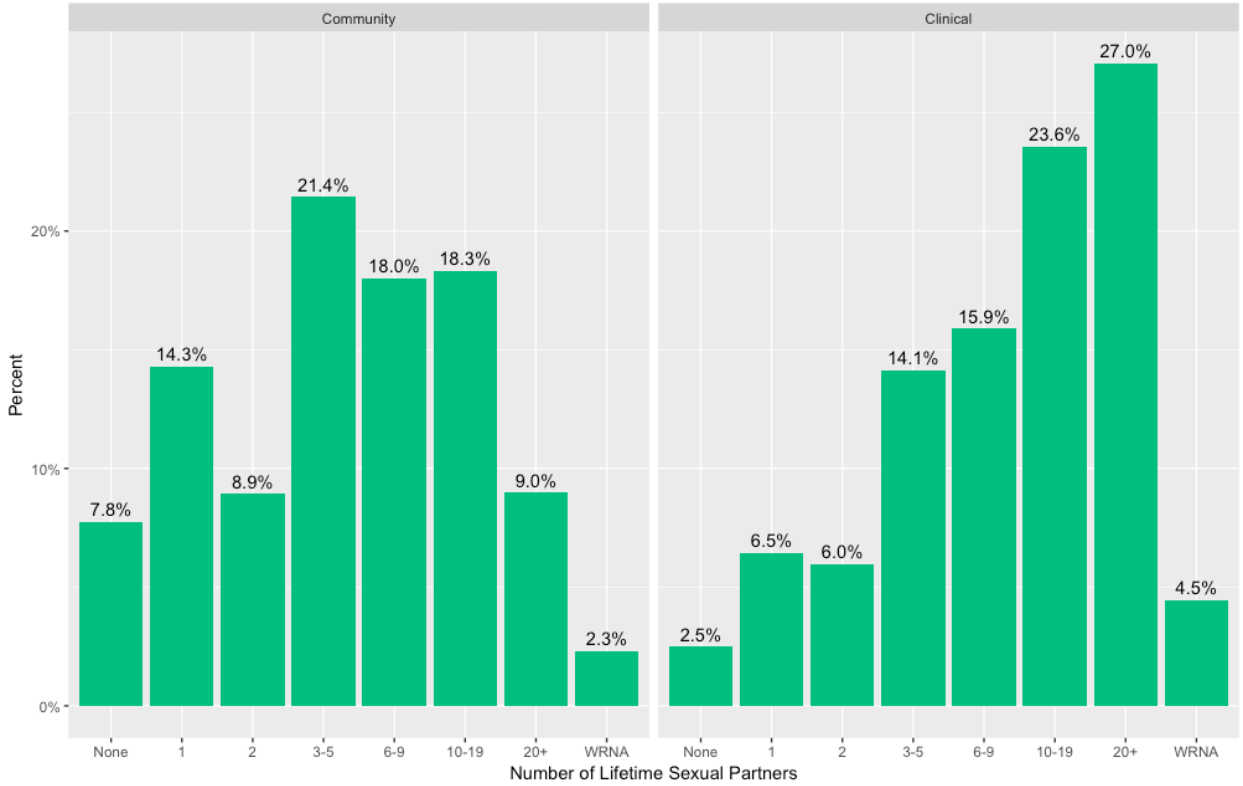
Though we found low variability rates of STI/STD diagnoses, it is possible that this item identifies individuals with extreme outcomes (i.e. via a follow up question that asks how many times a participants has been told they have a particular diagnosis). Repeated diagnoses were rare (among those with a diagnosis, 72.5%-100% of diseases or infections occurred a single time). The most common reoccurring disorder by percent (72.5%) of diagnoses was gonorrhea (e.g. 11 out of 40 individuals reported more than one diagnosis). However, it is not certain if participants were reporting on repeated exposures/ reinfections or multiple consultations for a single exposure.

Unfortunately, the M-RBQ did not assess unwanted pregnancy. Past 12 month pregnancy was assessed using the Life Experiences Survey (Sarason, Johnson, & Siegel, 1978). While 8.6% of the responding sample (n=393) indicated that they or a spouse became pregnant in the past year, it was unclear whether these pregnancies were intended. This measure also assessed past year abortion, which was endorsed by only 1.0% of the responding sample (n=47). Thus, associations with potential RSB measures and this negative health outcome could not be assessed.

Number of lifetime sexual partners

Figure 4.2.

Frequency of responses for number of lifetime sexual partners



Note: “Would rather not answer” = WRNA

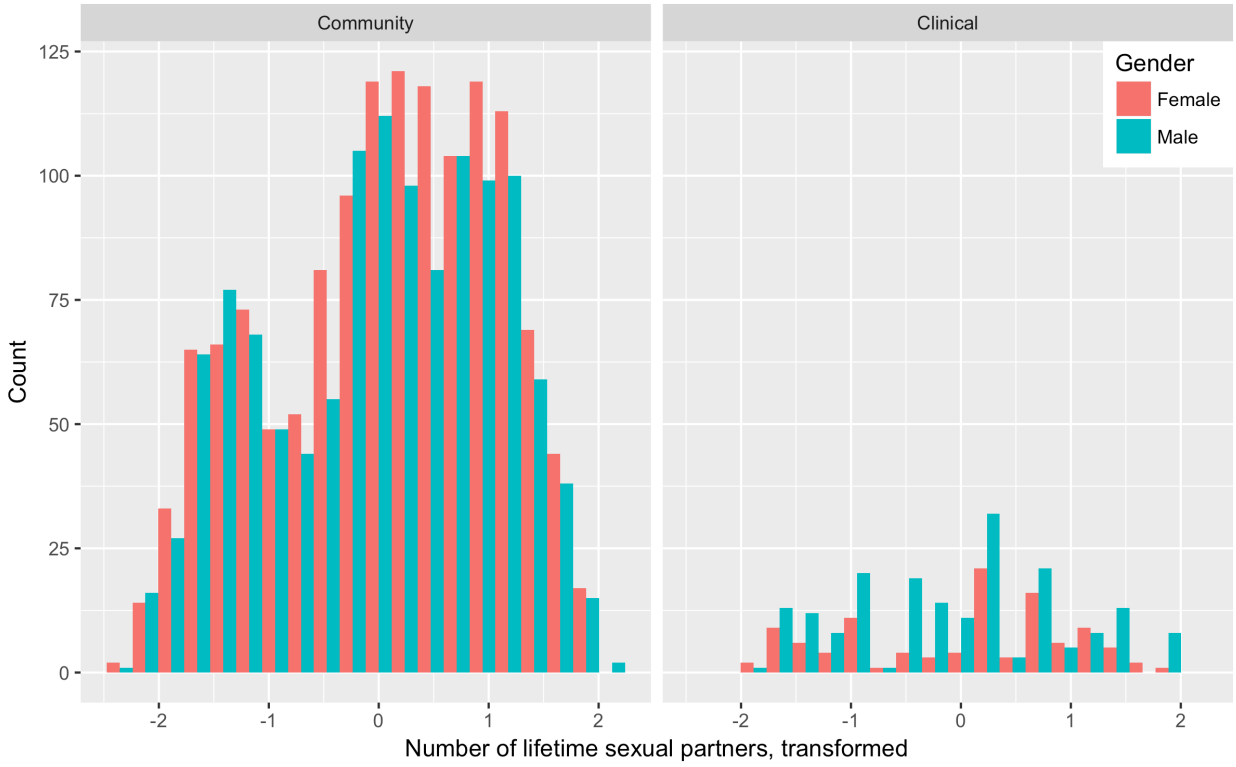
The number of lifetime sexual partners variable had the highest number of informative responses (n=3,758) compared to all other variables because it was the first of two major skip out items.

Unlike previous studies, we were unable to calculate a mean number of partners due to the categorical response options. However, several trends of the data were consistent with what would be expected (i.e. males typically reported higher number of lifetime sexual partners compared to females, and the clinical sample typically reported higher number of lifetime sexual compared to the community sample).

To account for gender, age, and clinical differences, we created a quasi-continuous variable (previously discussed in Chapter III). Using the community sample, scores were created using standard regression procedures and then z-scored within sex. Clinical participants were then z-scored using the mean and standard deviation of the community sample. As the scores were standardized on the community sample (m= 0, $s.d.$= 1 for males and females), we do see elevated scores for clinical males (m= 0.345, $s.d.$= 1.06) and females (m= 0.353, $s.d.$= 0.93). These estimates are based on a slightly smaller sample (n=2,931) used in Chapter III (See Figure 4.3).

Figure 4.3.

Distribution of responses for number of lifetime sexual partners, transformed



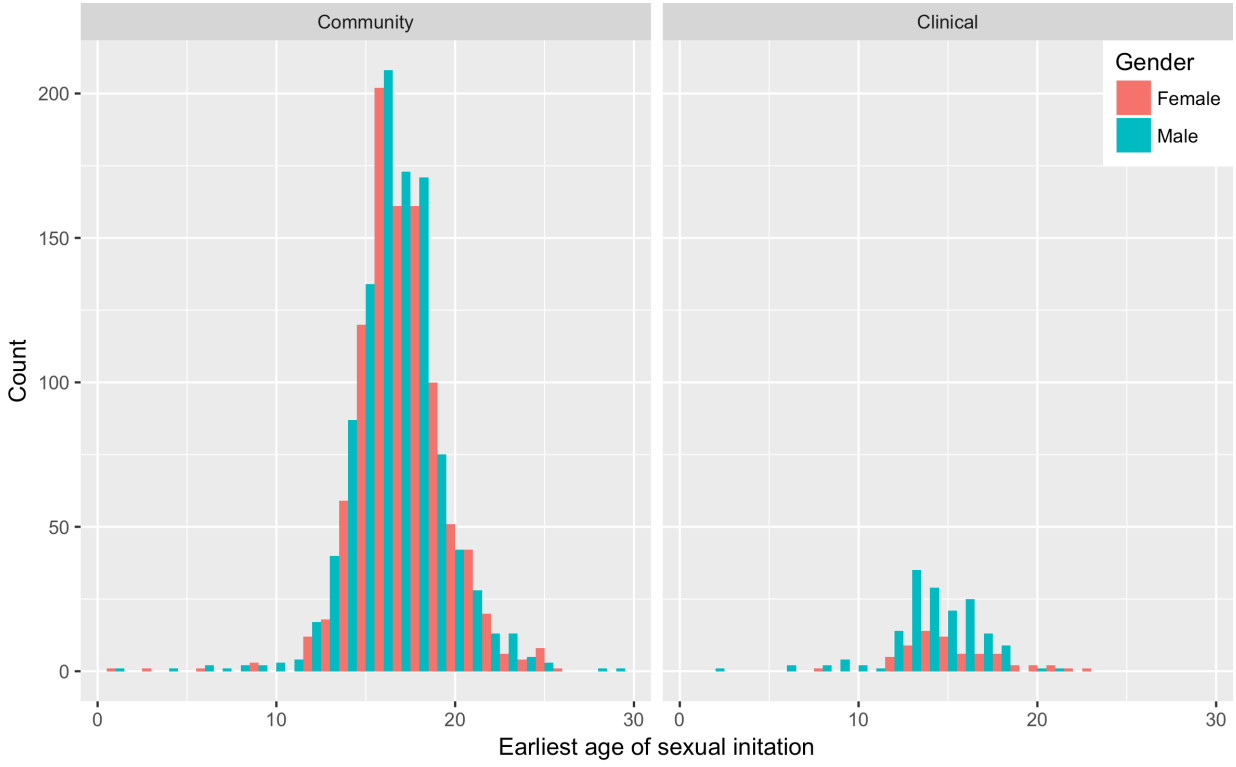
Note: Figure portrays the original sample in which transformed scores were calculated (n=2,931).

Age of initiation

Due to the two skip out items, scores for earliest age of sexual initiation (across anal, oral, or vaginal sex) were only available for participants who had more than one sexual partner in the past five years (e.g., the M-RBQ sample). In the third wave of CADD data collection, scores for age of sexual initiation were available for 1,725 participants. The distribution of age of sexual initiation was fairly normal. The mean age for the community sample was consistent with previous reports (*m*=17.11, *s.d.*= 0.50 for females, *m*=16.67, *s.d*=0.49 for males), though we see earlier average age of sexual initiation in the clinical sample (*m*=15.40, *s.d.*= 0.50 for females, *m*=14.16, *s.d.*=0.49 for males).

Figure 4.4

Distribution of responses for earliest age of sexual initation for anal, oral, or vaginal sex



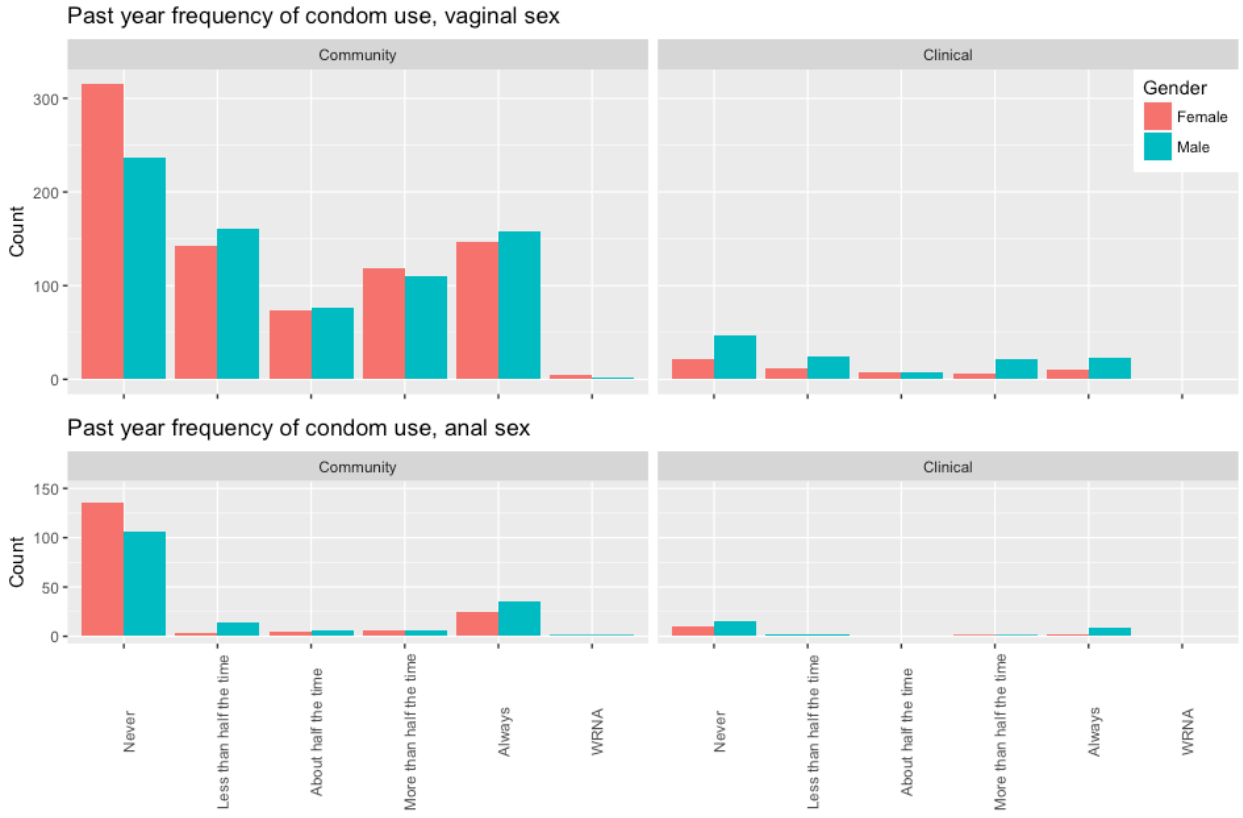
Frequency of condom use

Condom use frequency over the past 12 months varied across gender, type of specific sex act, and gender of partner.

Heterosexual sex was the most common in this sample, though condom use rates varied by act (See Figure 4.5). 19.9% of males who have vaginal sex with females report always using condoms, while 17.6% of females reported their male partners always use condoms during vaginal sex. For heterosexual anal sex, 21.4% of males who have anal sex with females report always using condoms, while 13.0% of females reported their male partners always use condoms during anal sex.

Figure 4.5.

Past year condom use for penetrative sex in heterosexual participants

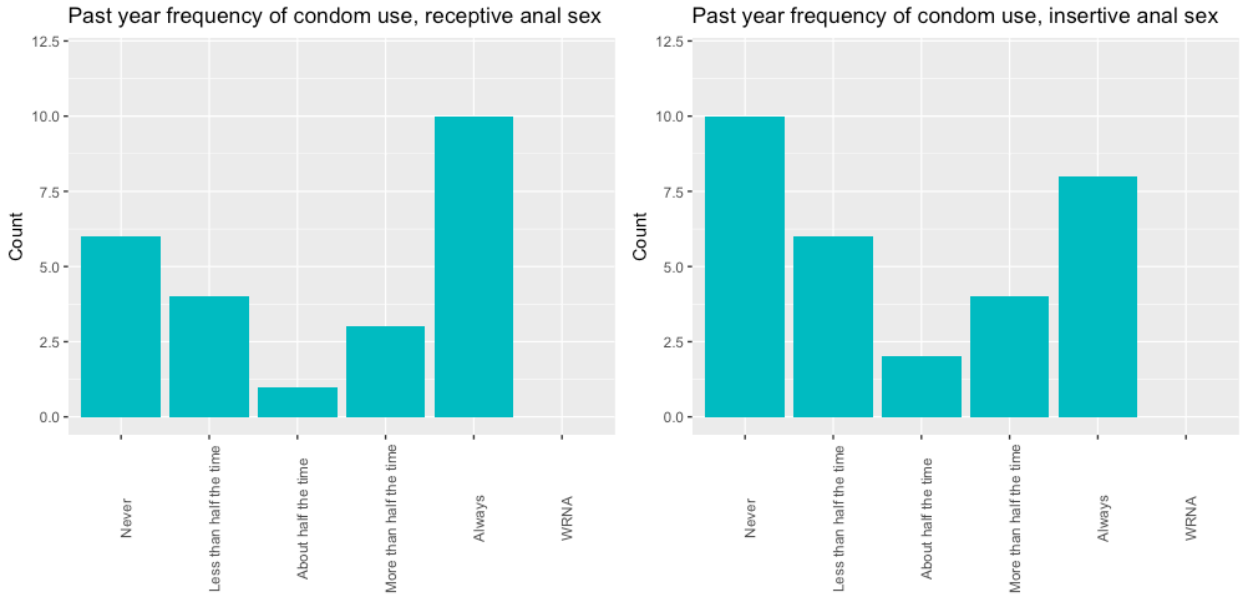


Note: "Would rather not answer" = WRNA

Consistent condom use was highest among males who have anal sex with males (See Figure 4.6). 26.3% of males report always using condoms during insertive anal sex, and 41.4% of males report always using condoms during receptive anal sex. While these percentages are high, this group is a small fraction of the overall sample.

Figure 4.6

Past year condom use during anal sex for men who have sex with men

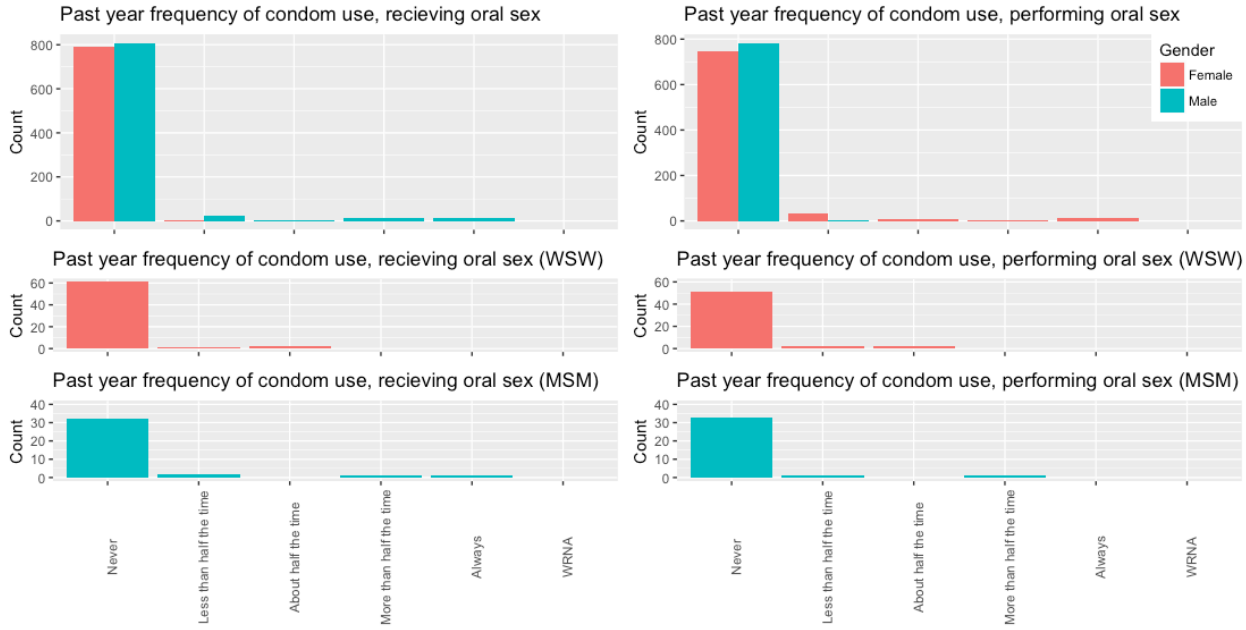


Note: “Would rather not answer” = WRNA, “MSM” = men who have sex with men

Consistent condom use or dental dam use was virtually nonexistent for oral sex between females or when males performed oral sex on females (See Figure 4.7). Oral sex between males had somewhat higher rates of consistent condom use, with 4.2% of oral sex recipients and 2.2% of oral sex performers reporting always using condoms.

Figure 4.7

Past year condom/ dental dam use during oral sex



Note: "Would rather not answer" = WRNA, "WSW" = women who have sex with women, "MSM" = men who have sex with men

There are several complexities to creating a sum-score of condom use frequency. In terms of potential exposures, frequency should be weighted by number of exposures (e.g. someone who reports rare condom use and rare sexual encounters could have fewer exposures than someone who is more active with slightly inconsistent condom use). It is difficult to combine such a statistic across specific sex acts, as it was unmeasured whether the various acts are with the same or different partners. Additionally, contextual factors such as partner type or characteristics would alter risk of acquiring STI/STDs.

Sex under the influence

Our measure of sex under the influence did not measure frequency of drug and alcohol use during sex, per se, but rather measured how drug and alcohol use may have impacted sexual decision making. The frequency of four past year behaviors were assessed, with most

participants reporting few occurrences of these behaviors (see Figure 4.8). Thus, we created a composite score to aggregate these behaviors into a single item (as described in Chapter III).

Figure 4.8

Past year frequency of four sex under the influence measures for M-RBQ sample (n= 2,318)

Note: "Would rather not answer" = WRNA, "Low risk" includes those with zero lifetime partners or zero or one partner(s) in the past five years

In addition to these items, the M-RBQ also include two items aimed at measuring whether drug and alcohol use was related to condomless sex. The first item, "were you using alcohol or drugs the last time you had unprotected sex and DID NOT use a condom for protection," assessed the percentage of inconsistent condom users who may have been made

more risky by using substances. Out of 2,253 responders, 27% reported unprotected sex was paired with substance use (n=606), 66% reported that their unprotected sex was not due to substance (i.e. occurred when sober, n=1497), and 7% reported that they never use condoms with or without substance use (n=150). The second item, "were you using alcohol or drugs the last time you had sex and DID use a condom?," assessed when there was an absence of risk behavior despite the use of substances. Out of 2,260 responders, 24% reported using condoms despite using substances (n=547), 62% reported being sober during last use of condom for protection (n=1409), and 14% reported that they always use condoms with or without substance use (n=304). While only a small percentage of participants who report that their condom use is stable both when sober and using substances, condom non use was not entirely linked to substance use during sexual decision making.

Risky sex acts

Rates of anal sex were investigated as the primary measure of high risk behavior, given the increased risk of STI/STD transmission. Our sample included only 76 males who reported having male sex partners for anal or oral sex (e.g., men who have sex with men). In this group, 48 men (63.2%) reported having insertive anal sex with a male partner and 44 (57.9%) reported having receptive anal sex with a male partner, with substantial overlap. We identified only four males who had had insertive anal sex with male partners but not receptive anal sex. Additionally, we identified only eight males who had receptive anal sex but not insertive anal sex.

Rates of heterosexual anal sex were similar to those reported in the National Survey of Family Growth (35.9% of women and 42.3% of men; Copen, Chandra, & Febo-Vasquez, 2016). Of the 1,047 females who reported having male sex partners or anal, oral, or vaginal sex, 408

(39.0%) reported having receptive anal sex. Out of 1,217 males who report having female partners, 520 (42.7%) reported having anal sex.

The second measured high risk behavior was trading sex for drugs, alcohol or money, which has a relatively low prevalence in the general public. In the AddHealth sample, approximately 3.5% of adolescents exchanged sex for drugs and/or money and also had higher rates of HIV or other STI/STDs (Edwards, Iritani, & Hallfors, 2006). Similarly, within the CADD samples few endorsed this behavior (See Table 4.2). Among participants who reported trading sex for drugs and alcohol (n=35) or money (n=23) at least once in their lifetime, past year high risk behavior was low (See Figures 4.9). Only 11 participants reported both behaviors in their lifetimes.

Table 4.2.

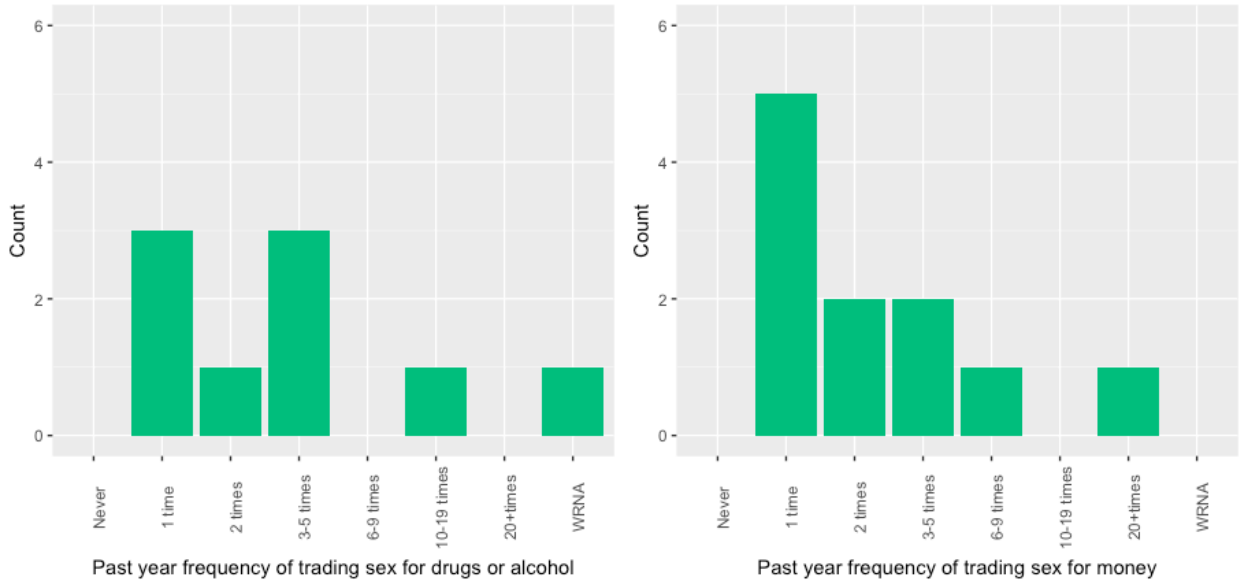
Percent of responses for lifetime high risk sexual behavior for M-RBQ sample (n= 2,318)

	Never	At least 1	WRNA
Given sex to get drugs	98.3	1.5	0.1
Given sex to get money	98.8	1.0	0.2

Note: "Would rather not answer" = WRNA

Figure 4.9.

Past year frequency of high risk behavior



Note: "Would rather not answer" = WRNA

Risky sexual partners

Information about partner type (i.e., regular [ongoing, romantic, or committed partnerships] and casual [e.g., non-romantic, or uncommitted partnerships]) was unavailable from the M-RBQ. However, there was a single item that assessed consensual or non-consensual partner non-concurrency.

Other partner characteristics that may confer risk include those with STI/STDs or HIV or those with high risk for these diseases (i.e. partners who are injection drug users or smoke crack). Participants reported on past 12 month partners (e.g. "In the last 12 months, how many of your sex partners were having sex with someone other than you?" or "In the last 12 months, how many of your sex partners had hepatitis?").

Partner concurrency was the most prevalent in the sample, with relatively low affirmative responses about STI/STDs or HIV and other high risk behavior (See Table 4.3)

Table 4.3.

Percent of responses for partner type variables for M-RBQ sample (n= 2,318)

	None	IDK	At least 1	WRNA
Other partners	66.4	13.6	19.3	0.7
Injection drug users	95.7	2.4	1.3	0.7
Crack smokers	96.7	1.7	1.0	0.5
HIV or AIDs	97.5	2.0	>0.1	0.5
Hepatitis	96.8	2.5	0.2	0.5
Gonorrhea or syphilis	96.7	2.4	0.5	0.4
HPV or NGU	92.0	3.1	4.5	0.4

Note: "I don't know" = IDK, "Would rather not answer" = WRNA

Across all behaviors and diagnoses, a small percentage of participants endorsed the "I don't know response" regarding their partner's behavior. Fewest participants were uncertain about partners potential crack cocaine use (n=40). Uncertainty about partner's STI/STD status was fairly similar across specific diagnoses, though slightly higher for HPV, which is a common and sometime symptomless infection. Uncertainty about partner concurrency was highest (n=310). Interestingly, uncertainty about partner's injection drug use was similar to the uncertainty about STI/STDs. Such responses could indicate 1) true ignorance about partner or 2) an acknowledgement of possible risk. The first case would suggest that more participants had conversations with their partners about STI/STDs and drug use than about monogamy, given differential rates of uncertainty. However, it is possible that more people are uncertain about their partner's monogamy than their drug use because they are making an estimate based on likelihood of behavior (i.e. are more likely to acknowledge that there is a possibility that they cannot truly know about these characteristics in their partners). Therefore, this response was ambiguous and not a clear indicator of level of risk. Finally, there is evidence that people are often inaccurate in predicting their partner's concurrency and STI/STD status (Drumright, Forbach, & Holms, 2004).

Other measures of RSB

Finally, the M-RBQ included some items that did not fall into the previously described categories of RSB. These measures included HIV expectations (i.e. participants expectation of acquiring HIV and frequency of HIV testing: See Figures 4.10- 4.11).

Figure 4.10.

Distribution of responses for participant's expectancy of acquiring HIV

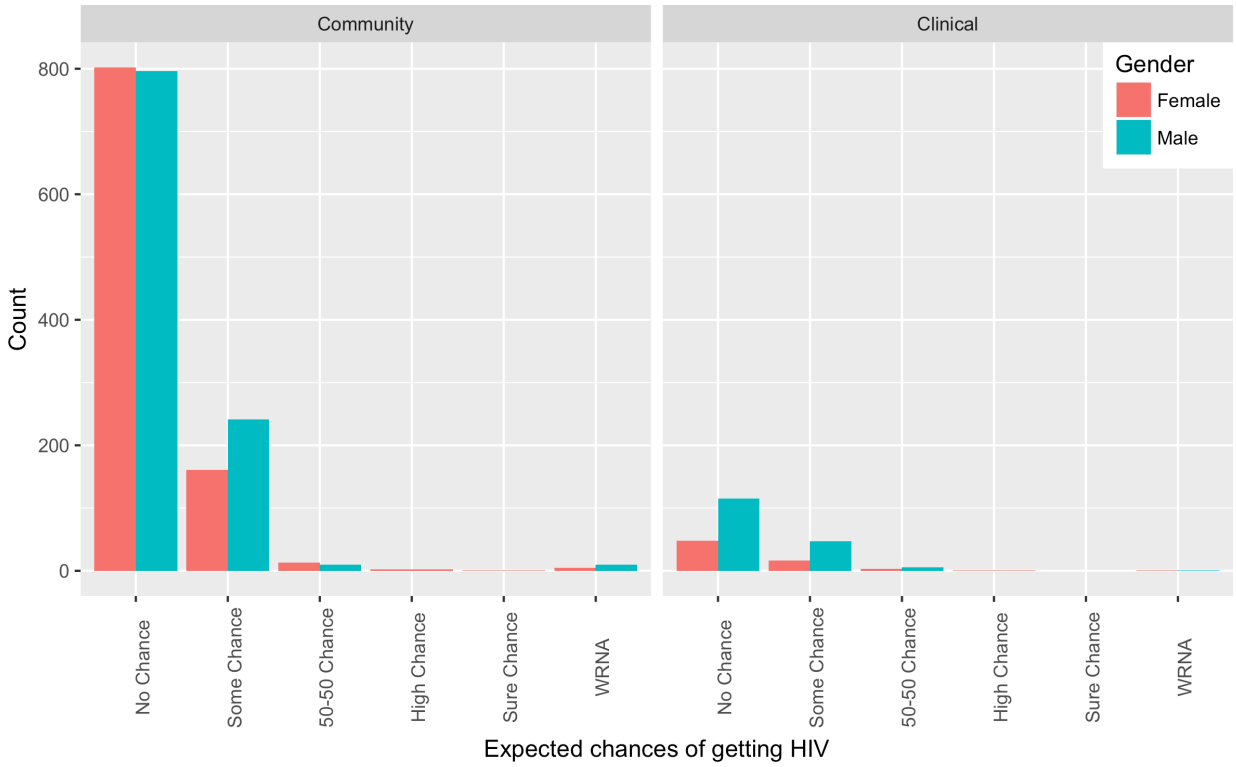
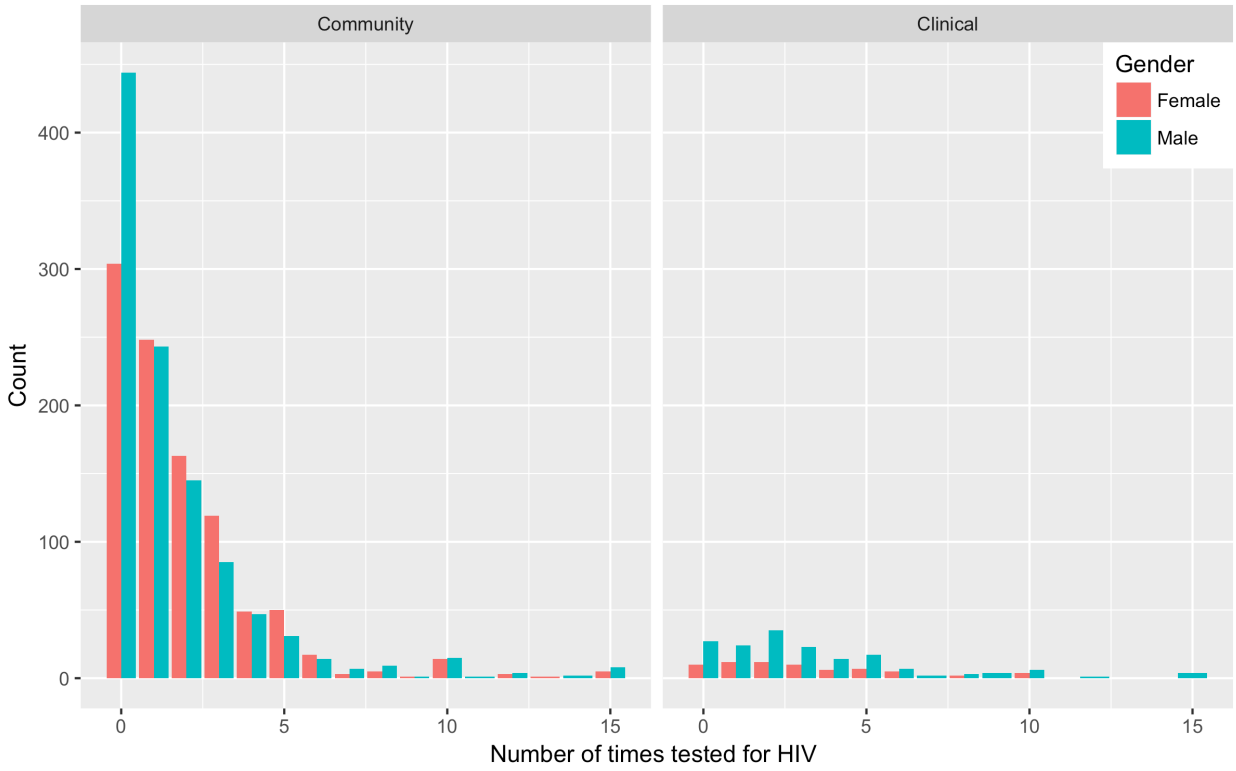


Figure 4.11.

Distribution of responses for number of times tested for HIV



Note: Scores were truncated at 15 tests.

These measures are not common in the literature, for several possible reasons. Though there is strong evidence (i.e. across 27 studies globally) that an individual's self-rated health is good predictor of mortality (Ilder & Benyamini, 1997); it is unclear whether individuals are also good predictors of their HIV risk. The measure of HIV testing was created under the assumption that those with risker behaviors may seek out more HIV tests because of a recognition of their increased susceptibility to HIV. However, it is possible that repeat HIV testing could also be a continual need for reassurance, a preventative measure for transmission or a risk reduction. Looking at differences between first time testers and repeat testers in a London HIV clinic, repeat testers did engage in more RSB in only a subset of the sample (i.e. for MSM, and no differences were found for heterosexual individuals; Leaity et al., 2000). Another study of MSM

found a similar association or RSB with repeat testing; however, the study also identified a group of regular testers who intentionally seek out tests (compared to a passive test such as during a blood donation, to qualify for health insurance, during routine physical exam, etc.), who rate themselves as having a lower average risk of HIV, and who were younger on average (Fernández et al., 2003). Thus, a simple count of number of HIV tests is not a direct predictor of negative health outcomes.

Biometric Models

We selected the most promising measures for genetic analyses, as a way to identify the most heritable measure with little sex limitation. Univariate biometric modeling for two of the variables, number of lifetime partners and sex under the influence was previously described in Chapter III. As age of sexual initiation met several criteria (e.g. continuous and normally distributed, not effected by low prevalence), we estimated heritability using the twin sample described in Chapter III. While this variable was moderately heritable (See Table 4.4), constraining path estimates across gender did significantly reduce the fit of the model ($\chi^2_{diff}(3)= 16.483$, $p<.01$), indicating sex-limitation. Age of initiation for females was influenced primarily by shared environmental influences. The low heritability among females suggests that age of sexual initiation may be an appropriate RSB measure for genetic analyses in males only. While we did not find evidence for sex limitation for sex under the influence, this variable had low heritability ($a^2= 19\%$). Thus, number of lifetime partners best fit the biometric criteria ($a^2= 50\%$).

Table 4.4

Standardized path loadings for canidate measures of RSB and twin correlations

Model	rMZF	rMZM	rDZF	rDZM	rDZOS		a_{11}^2	c_{11}^2	e_{11}^2
1) Age of sexual initiation						**Female**	**0.13**	**0.40**	**0.47**
						Male	**0.56**	**0.01**	**0.43**
	.529	.530	.431	.172	.240	Constrained	0.51	0.03	0.45
2) Sex under the influence						**Female**	**0.02**	**0.32**	**0.65**
						Male	**0.27**	**0.07**	**0.65**
	.347	.345	.334	.212	.193	Constrained	0.13	0.21	0.65
3) Number of lifetime sexual partners						Female	0.42	0.15	0.44
						Male	0.53	0.11	0.36
	.661	.655	.335	.365	.312	**Constrained**	**0.46**	**0.13**	**0.40**

Note. Female and male paths estimated separately are compared to a constrained model with paths estimated jointly. **Bold** indicates the best fitting model.

Association with risk

Though not a criteria of a candidate variable, we hypothesized that any measure of RSB should be correlated with other indicators of BD. Given the clinical sample was selected for elevated rates of behavioral disinhibition, we expect to see the clinical sample with elevated rates of any RSB measure. There were significant mean differences between the community and clinical samples for lifetime number of sexual partners and age of sexual initiation. There was limited power to detect differences in low prevalence variables and negative health outcomes (e.g., STI/STD diagnoses, sex under the influence).

Pilot study

Table 4.5 describes the distribution of age, sex, and self-reported ethnicity in the pilot sample. Notably, this sample was younger than the CADD study, which may bias at least two of the measures of RSB. As number of lifetime sexual partners is cumulative, we expect younger samples to have fewer lifetime partners on average compared to older samples. That is the case here, when comparing this college sample to the CADD community sample. Additionally, we

would expect a portion of the sample to have not yet initiated sexual activity. Approximately 20.33% of the participants (n=37) had not yet indicated anal, oral, or vaginal sex at the time of the study.

Table 4.5

Undergraduate sample descriptives and primary measures of RSB

	All (N=182)	Male (N=65)	Female (N=117)
Mean age in years	18.73 (1.22)	18.97 (1.25)	18.60 (1.19)
Self reported ethnicity			
White, Caucasian, Euro-American	70.33%	66.15%	72.65%
Asian or Asian American	10.44%	4.62%	13.68%
Latino or Hispanic, non-Black	7.69%	13.85%	4.27%
Biracial or Multiracial	6.04%	6.15%	5.98%
Black	2.20%	1.54%	2.56%
Middle Eastern	2.20%	4.62%	0.85%
Pacific Islander	1.10%	3.08%	0.00%
Number of lifetime sexual partners			
None	20.33%	20.00%	20.51%
1	17.58%	13.85%	19.66%
2	16.48%	18.46%	15.38%
3-5	20.88%	23.08%	19.66%
6-9	16.48%	16.92%	16.24%
10-19	6.59%	6.15%	6.84%
20+	1.10%	1.54%	0.85%
Mean age of sexual initiation in years			
Anal	17.50 (1.72)	18.00 (1.50)	17.20 (1.87)
Oral	16.01 (1.52)	15.94 (1.53)	16.06 (1.51)
Vaginal	16.52 (1.38)	16.92 (1.51)	16.28 (1.48)

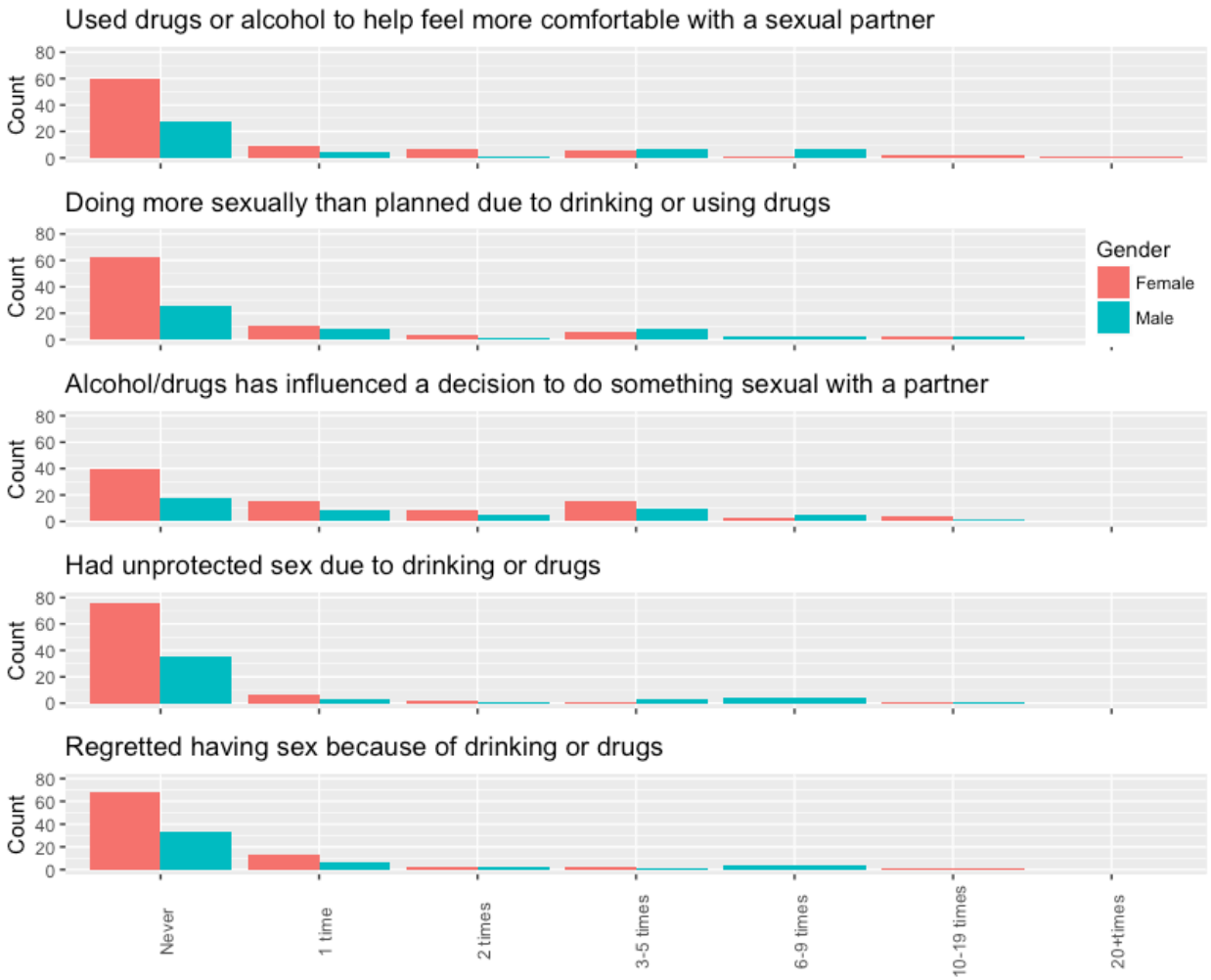
Note: Raw scores indicated means, parentheses indicate standard deviations (*s.d.).*

The primary measure of the pilot study was testing the four items measuring sex under the influence from the M-RBQ and an additional item assessing regret of a sexual encounter due

to alcohol or drugs. Frequency of these behaviors were low in this sample (See Figure 4.12). The most frequently endorsed behavior was that drugs or alcohol influenced a decision to do something sexual with a partner.

Figure 4.12.

Past year frequency of five sex under the influence measures for undergraduate sample



We were interested in testing gender effects on these behaviors, as well as gender effects in motivations for endorsing these variables. Collapsing responses into never vs. any sex under the influence behavior reported, there were no gender differences for using drugs and alcohol to feel more comfortable with a partner ($\chi^2(1)= 0.01$, $p=0.91$), doing more than planned because of being drunk or high ($\chi^2(1)= 1.22$, $p=0.27$), that drugs or alcohol influenced sexual decision

making ($\chi^2(1)= 0.00$, p=0.98), or having unprotected sex because of being drunk or high ($\chi^2(1)= 0.21$, p=0.65). Inclusion of a new item that measured whether or not participants regretted having sex because of drinking or using drugs showed a marginally significant gender trend, where females were slightly more likely to report regretting a sexual encounter ($\chi^2(1)= 3.00$, p=0.08).

Though we did not find gender differences for these items, we tested for other possible differences between men and women in the sample. We found no significant differences between men and women on number of lifetime partners ($\chi^2(6)= 1.49$, p=.96), or past year number of sexual partners ($\chi^2(6)= 6.84$, p=0.34). There were no significant mean differences for age of sexual initiation across all types of specific sex acts (e.g. anal [$t(21)= 1.08$, p=0.743)], oral [$t(136)= 7.19$, p=0.270)], or vaginal sex [$t(126)= 2.34$, p=0.271)]). Similarly, there were no gender differences in other recreational risky behaviors including 1) rock climbing, sky diving, or hang gliding ($\chi^2(4)= 2.29$, p=0.68), 2), skiing, snowboarding or skateboarding ($\chi^2(4)= 6.82$, p=0.15), 3) water skiing, surfing, or SCUBA ($\chi^2(4)= 5.92$, p=0.21), 4) cycling, mountain biking, or BMX ($\chi^2(4)= 7.80$ p=0.10, 5) snowmobiling, jet skiing, or boat racing ($\chi^2(4)= 4.66$, p=0.32), and 6) riding motorcycles, dirt bikes or ATVs ($\chi^2(4)= 5.85$ p=0.21). Our questionnaire also assessed drug use, and found no gender differences for general marijuana use ($\chi^2(1)= 0.12$, p=0.73) or frequency of marijuana use before driving ($\chi^2(6)= 6.72$, p=0.35). The only significant gender effect identified was that males reported more frequent marijuana use before work ($\chi^2(6)= 16.64$, p=0.01).

Finally, responses from the sex under the influence follow up variables show a variety of motivations for endorsing these items. While gender differences were not explicitly tested due to lower power and multiple testing issues, some gender trends are discussed below.

The most common sex under the influence behavior endorsed was that alcohol or drugs influenced a sexual decision (n=101). A small percentage of participants, 1% of males (n=3) and 22% of female (n=14) endorsers report making a sexual decision that they would not have chosen if sober. However, 57% of male (n=20) and 26% of female (n=17) endorsers admitted that their decision would have likely been the same if sober at the time of the sexual encounter (e.g., they happened to be high or drunk at the time of the decision, but the decision was not actually modified by substance used). An additional 29% of males (n=10) and 40% of female (n=26) endorsers say a sexual decision was easier to make because of drugs or alcohol, suggesting possible effects of using drugs and alcohol as a social lubricant (e.g., substance use helped to solidify their decision). For the remaining open-ended responses, most females reported that they were feeling "more confident" because of drug or alcohol use, while males reported that drug use 1) "intensified pleasure," 2) made them "too high to initiate but not to resist," or 3) "too high and uncomfortable to have sex." Thus, these comments indicate that endorsing this behavior does not necessarily reflect that that sex was made risker from drug and alcohol use.

Endorsement of using drugs or alcohol to feel more comfortable with a partner was more straightforward. Almost 40% of endorsers said that they used drugs or alcohol to feel less shy around a partner (n=29), followed by 23% who used to feel more turned on (n=17), 10% who used to reduce physical discomfort associated with sex (n=7), 10% who used to be less shy about specific requests or desires (n=7), and 8% who used to feel more comfortable about a sexual decision (n=6). There were few gender differences, except that zero males needed to use drugs or alcohol to feel comfortable expressing requests or desires. Males added that they used drugs or

alcohol to “last longer” and to have a “new experience.” Females also reported that drugs and alcohol “changed things up.”

Similarly, there were few gender differences for reasons for regretting sex under the influence. About half of endorsers regretted sex because they would not have had sex with that person or at that time sober, with 53% (n=8) of males and 51% (n=19) of females selecting this option. Alternatively, 40% (n=6) of males and 49% (n=18) of females who regretted sex under the influence did so because it resulted in social consequences. Only one male reported regret due to a physical consequence (e.g. new STD or contributed to an unintended pregnancy).

There were more gendered trends for responses regarding the frequency of doing “more than planned” with a sexual partner because of drugs or alcohol, as well as a more complicated pattern of results. It is possible that this item reflects that sex was made risker from drug and alcohol use, especially if the person was not initially planning on having sexual contact with that partner or within a specific situation. Nearly half of the participants (n=22) reported this was the case, though 83% (n=5) of males and only 31% (n=8) of females reported being happy with the outcome. The remaining participants (n=36) reported that they were hoping to have sexual contact, but went further than planned because they were drunk or high. Similarly, 88% (n=14) of males and only 55% (n=11) females were satisfied with this outcome. One female reported being “taken advantage of” and another female reported regret because the act “affected a friendship negatively.”

While participants reported on frequency of having unprotected sex *due to* being drunk or high, the pattern of suggests that some responses are measuring unprotected sex that is co-occurring with drug and alcohol use. In terms of drugs and alcohol making sex more risky, about 50% (n=6) of males and 62% (n=18) of females reported any instance of having unprotected sex

while drunk or high that would have been protected if sober (e.g. because it was to difficult to get a condom or they forgot). The remaining 50% (n=6) of males and 38% (n=11) of females reported that they did have unprotected sex while drunk or high but that that sex would have been unprotected if sober for a variety of reasons (e.g. do not like the feeling of condoms, do not have heterosexual vaginal sex, pregnancy concerns, or use other forms of birth control). Finally, one male reported not using condoms but looks for "signs for disease."

Discussion

Given issues regarding the measurement of sexual behavior outlined above, this chapter included two exploratory investigations. Information gathered from the pilot study informed the interpretation of RSB variables (i.e. primarily sex under the influence) derived from our risky sexual behavior inventory (i.e., M-RBQ). Suggestions from the literature, as well as results from the pilot study and the CADD sample, were considered in the revision of the M-RBQ. Finally, these analyses were used to select an optimal phenotype to be used in later genetic analyses (i.e., in Chapter IV).

Changes to the M-RBQ

The motivation behind changes to the M-RBQ directly addressed suggestions in the literature and feedback from the pilot study, all while serving the conflicting goal of creating a paired down, efficient version of the instrument. Final changes were approved by a committee of faculty at the Institute for Behavioral Genetics. Changes were made in the following ways (for the final version of the Short M-RBQ, see Appendix A).

A major limitation of the M-RBQ was the skip out structure, which created a large amount of missing data. Participants with zero lifetime sexual partners (n=278, 7.2% of participants in the CADD sample) did not complete the questionnaire. Additionally, a second

skip out was included which relied on the assumption that participants with zero or one sex partners in the past five years were not engaging in risky sexual behavior (e.g. response to item "In the past five years, with how many different people have you had vaginal, oral, or anal sex with?"). While a majority of the questionnaire would be inappropriate for those reporting zero partners in the past year (n= 94, 2.6% of sexually active participants in the CADD sample), several lifetime measures such as age of sexual initiation, or lifetime STI/STD information would have been informative for those reporting a single partner (n=1,087, 30.4% of sexually active participants) in the past five years. Though there was an assumption of low RSB in recent year, it was possible that those with one partner could have recent risky behaviors such as sex under the influence, sex with a risky partner, or indicate previous RSB on the lifetime measures. Thus, inclusion of this skip out not only reduced our base rates of both risk behavior but also likely reduced base rates of negative health outcomes such as lifetime STI/STD diagnoses. As this skip out excluded a large proportion of the total sample, it was removed in the final revision.

In contrast to other surveys, many of the questions were nested within a specific sex act including age of onset and past year number of partners (see Figure 4.1). Though designed to provide detailed information about each specific sex act, this structure presented several limitations. While it was possible to calculate age of onset across specific sex acts (i.e., a common measure used in the literature) by taking the minimum reported onset, other variables such as number of past year partners or frequency of condom use was difficult to combine. For example, past year number of partners could not simply be combined across variables as the same partners were likely counted across multiple specific sex acts. Further, the ordinal scale was not additive and would not result in a true count. To better assess this item in the new

version, “in the past 12 months, with how many different people (different sex partners) have you had oral, vaginal, or anal sex?” was added to the final version.

Additionally, the follow up questions for specific sex acts were considered to be tedious for many active subjects. For instance, an average respondent who endorsed having vaginal sex and performing/receiving oral sex would be asked to answer 18 follow up items (e.g., six items per act, see Figure 4.1). In an extreme case, a participant could be asked up to 42 sex act specific items if they had endorsed having each specific sex act with both male and female partners. Given that these items were not predictive of negative health outcomes or informative of RSB (i.e. the association or direction of risk could not be determined), these items were reduced in the final version.

Furthermore, we removed an item attempting to measure frequency of STI/STDs. Previously, if a participant answered that they had ever been told by a doctor, nurse, or health care professional about an STI/STD diagnosis, they were prompted to indicate the number of times. While the base rates of STI/STDs in our sample was low (ranging from 0-6.8% for any given diagnosis (See Table 4.1) and likely biased by the skip-out, the percent reporting multiple times was very minimal. It was unclear whether participants reporting two or more times for other diagnoses (e.g., hepatitis C, syphilis, gonorrhea, or chlamydia) were reporting on repeated exposures/ reinfection or multiple consultations for the same exposure.

Several other standalone questions were uninformative of RSB or outcomes, due to either ambiguous wording of the questions or ambiguous response items. As previously mentioned, an “I don’t know” option was available for all items. In the case of items assessing risk based on characteristics of sex partners (e.g., in the past 12 months “how many of your sex partners were having sex with someone other than you?, ”“how many of your sex partners had hepatitis,” etc.),

this option was fairly common. While it thought to reflect sexual risk with uncommitted or unknown partners for participants ignoring the potential risk, it is also possible that conscientious participants are acknowledging the realistic possibility that there is always a risk for uncertainty in partners' behavior. The question "In the past 12 months, have you had sex with someone other than your steady partner while still in the relationship," was not informative of risk without additional contextual information such as partner type, safe sex precautions, etc. and was removed from the final version. Wording was modified (e.g. "partner "was changed to "sex partner") to improve clarity of some items (Mercer, 2010).

To better address casual sex, we included several additional items from a *sexual risk taking with uncommitted partners* factor on the Sexual Risk Survey (Turchik & Garske, 2009). These items included past year frequency for "times have you had sex with someone you didn't know well or just met" and "times have you had sex with a new partner before discussing their sexual history and risk for sexually transmitted diseases (STDs)." Participants were also asked about their past year partners, and prompted to report "how many of these sex partner would you consider a regular partner (with whom you have, or had a relationship)" and "how many of these sex partners would you consider a casual partner (with whom you did not have a relationship)." Additionally, the questions "in the last 12 months, how many sex partners (that you know of) have you had sex with who had been sexually active before you but had not been tested for sexually transmitted diseases (STDs) and HIV?" and "in the last 12 months, how many sex partners have you had sex with that you didn't trust?" were included.

Given the differential motivations to use condoms with regular vs. causal partners (Anderson et al., 1999), a follow up question about relationship type was added to the original condom use questions. For example, if a participant would also be asked "during those times

when you did not use condoms, were you in a regular (exclusive) relationship?" if they reported inconsistent condom use for vaginal or anal sex in the past 12 months.

To improve the utility of the questionnaire, several RSB outcomes were added in the revised format. As mentioned above, the original M-RBQ did not include a measure of unwanted pregnancy. Similarly, other studies have included unintended pregnancy as a proxy for unprotected sex (Huibregtse et al., 2011). In the revision, females are asked "have you ever had an unintended pregnancy?" and males are asked "have you ever gotten a partner pregnant unintentionally?," followed by "how many times" if the initial answer is "yes." Additionally, the item "in the past 12 months, how many of your sex partners had herpes?" was added to create a more complete list of STI/STDs.

To improve the sex under the influence item (previously described in Chapter III), the following item was included "in the past 12 months, I have regretted having sex with someone because I was using drugs or alcohol." This item was included to address differential social consequences for males and females for sexual behavior. Though there were not significant gender differences in our small pilot study, results suggested that this variable should be interpreted with caution. Unlike other assessments, these indicators do not measure the frequency that sex was made risker from drug and alcohol use directly. Rather, these questions may measure a more complicated phenotype (i.e., the extent to which drug and alcohol use co-occurred with sexual decision making), for which the motivation for engaging in or endorsing these behaviors may have gender effects.

We included a screener for possible sexual abuse, which was previously assessed in the CADD via the Colorado Adolescent Rearing Inventory (CARI-Q; Crowley et al., 2003) in only the early waves of data collection. For each specific sex act, participants were prompted to

corroborate that the age of initiation provided was the first time the subject "willingly agreed to" the specific sex act. This improved upon a previous proxy of abuse (e.g. age of partner during participant's specific sex act initiation), for which there is no clear age discrepancy cut off.

To improve inclusivity we included the following open ended item with a text box for the response, "Is there anything else we should know that may help us understand your responses? For example, were you unable to have sex during the past year due to an illness or injury? Or, due to military service, or your job, were you away from your sex partner for much of the past year?"

Finally, we built in pull down response options (e.g. a tab for age rather than an open text box) to reduce suspected typos that were common for measures of age of sexual initiation.

Optimal Phenotype for Genetic Analysis

Based on the criteria listed above (i.e., 1] moderate to high heritability, 2] reliable with low measurement error (i.e. easily standardized across samples), 3] little heterogeneity or sex-limitation, and 4] a continuous and normal distribution), we concluded that number of lifetime sexual partners is the best phenotype for use in genetic analyses.

Number of lifetime sexual partners was the most heritable phenotype of those explored in our sample (i.e. number of lifetime sexual partners, age of sexual initiation, and a composite variable of sex under the influence). Furthermore, our heritability estimate of 46% is consistent with moderate to high heritability estimates in other twin and family samples (Cherkas et al., 2004, Lyons et al., 2004, Mutanski et al., 2007, & Guo, Ton, Xie, & Lang, 2007). After controlling for mean differences across age and gender, this variable could be treated as a quasi-continuous variable, was fairly normally distributed, and had no evidence of sex-limitation.

Age of sexual initiation also met a fair amount of the criteria, with a few caveats. While age of sexual initiation was also quasi-continuous, normally distributed, and moderately heritable in our sample, there was evidence of sex-limitation. Though we found heritability to be much higher in males compared to females, previous twin and family studies have reported a wider range of heritability estimates for age of sexual initiation with inconsistent patterns of sex limitation (Harden 2014). For instance, Dunne et al. found strong social cohort effects moderated heritability of age of sexual initiation in a sample of Australian twins, as twins who "came of age" in the early 1970s and mid 1980s were influenced by moderate to strong genetic effects (a^2=.49 for females, and a^2=.72 for males) and twins who "came of age" prior to 1970 were less influenced by their genes (a^2=.32 for females, and a^2=.00 for males; 1997).

There were also several concerns about the reporting of age of sexual initiation that are specific to the M-RBQ in the CADD sample. Due to the skip-out, a large percentage of our sample was not asked this item. This is clearly not optimal for genetic analysis in Chapter V, due to the extent of missing data and resulting limited power. Only a limited amount of missing data could be recovered from earlier waves. Additionally, there were some outliers suggestive of very early sexual initiation (i.e.., n=30 participants reporting sex at 10 years of age or younger) that were very concerning. Given that participants are typing age into a response box, it is possible some of these extreme responses are typos (e.g. a participant types "8" instead of "18"; Indeed, several of these typos were verified when checking against reports in earlier waves of data collection) . It is also possible that participants are reporting on non-consensual or abusive acts. These responses could be verified with reports from earlier waves, reports on previous abuse and measured by CARI-Q (Crowley et al., 2003), or simply winsorized; the ambiguity of what constitutes very early sexual initiation (consensual vs. non-consensual) raises another valid

point. Though sexual initiation in adolescence is normative, legal age of consent in many states varies widely.

Finally, while it is common to combine data to define age of sexual initiation as earliest age of anal, oral, or vaginal sex, it is possible that the timing of specific sex acts, or the spacing and sequencing of other sex acts may be important. As such, recent efforts to categorize different patterns of sexual initiation have identified differential risk negative health outcomes such as STI/STD diagnoses (Haydon et al., 2012; Vasilenko, Kugler, Butera, & Lanza, 2015; Vasilenko, Kugler, & Lanza, 2015). For example, Add Health participants who initiated vaginal sex and other sex acts (i.e. oral and/or anal) within the same year had much higher odds of later STI/STD diagnoses and concurrent partners in adulthood compared to those initiated vaginal sex but delayed initiation of other sex acts (Haydon, Herring, & Halpern, 2012).

Other measures (e.g. inconsistent condom use, sex under the influence, risky sexual partners and acts, etc.) met even fewer criteria. Heritability estimates could not be calculated for many of these variables due to extremely low prevalence or ambiguous responses (i.e. "I don't know" responses about potential risky partners). Though overall prevalence for these behaviors was low, these items were informative about extreme risk behavior of a few individuals. For example, while an overwhelming majority of the sample did not trade sex for drugs, alcohol, or money, we did identify a single participant who reported 20 or more such transactions within the past year. Similarly, while most participants did not willingly have partners who were injection drug uses, eight participants reported multiple partners with this behavior including one participant reporting 20 or more injection drug using partners in the past year. A sum score adding risk indicators across different domains was theoretically possible, but there was little suggestion in the literature of how to properly weight this variable (e.g. what is the relative risk

of inconsistent condom use to knowingly having sex with a partner with HIV?). There are further concerns that a sum score would create mean differences in risk scores for large groups within our sample. For instance, though anal sex is relatively prevalent across the sample, the riskiest behavior in terms of STI/STD transmission is receptive anal sex (CDC, 2017). However, scoring receptive anal sex as an additional indicator of RSB may overweight females and men who have sex with men as risky rather than capture variation in the general population (e.g. heterosexual males and women who have sex with women). This may also create sex-limitation. Finally, we predicted that this would create a highly skewed zero-inflated ordinal variable, which would be inappropriate for further genetic analysis.

In conclusion, this chapter investigated issues in defining and measuring RSB. For further genetic analysis of an RSB in our samples, we identified the number of lifetime sexual partners to be the best candidate phenotype.

CHAPTER V

EVALUATING NUMBER OF LIFETIME SEXUAL PARTNERS WITH GENOME WIDE METHODS

Background

Number of lifetime sexual partners has been shown to be moderately to highly heritable in twin and family studies; estimates from the FinnTwin16 project were 55% and 42% for males and females respectively (Mustanski et al., 2007), whereas the estimate for females in the TwinUK sample was 41% (Cherkas et al., 2004). The results described in Chapter III show a similar pattern, where heritability is estimated at 61% and 42% for males and females respectively. When constraining the estimate across males and females, heritability was approximately 59% in the Australian Twin Registry. A related measure of multiple partners (i.e., categorical measure defined as ever having 10 or more partners in a single year) was 49% heritable in men (Lyons et al., 2004). Finally, estimates of heritability were consistent in females (50%) in the Add Health sample, though slightly lower for males (38%; Guo, Tong, Xie, and Lange, 2007).

Recently, genome wide approaches have begun to explore phenotypes related to sexual behavior and fitness. In a sample of 125,667 UK Biobank participants, Day et al. identified 38 genome wide significant loci associated with age of sexual initiation (2016). Recently, nine loci were associated with age at first birth (n=251,151) and three loci were associated with number of children ever born (n=343,072) in a large scale meta-analysis of 62 cohorts (Barban et al., 2016). However, no study has focused exclusively on number of lifetime sexual partners, a major indicator of RSB (see Chapter IV). Such methods are useful in that they 1) do not share the same assumptions as twin and family methods, 2) can be conducted on traditionally "unrelated"

samples, and 3) provide insight into potential causal genes and underlying biological mechanisms of a trait.

Genome wide association studies (GWAS) are used to identify loci associated with a phenotype of interest, typically single base pair differences in the genome (i.e. single nucleotide polymorphisms or SNPs). Associated SNPs tag or mark genetic variation nearby on the chromosome that is assumed to have causal effects on the phenotype, which identify target chromosomal regions for functional studies. Additionally, genome wide data can be used to estimate heritability directly from measured genetic variants (e.g. traditionally heritability has been estimated through twin and family studies). These estimates are derived from comparing the genetic similarity based on common SNPS in unrelated individuals to their phenotypic similarity using restricted maximum likelihood (REML) or genomic-related-matrix restricted maximum likelihood (GREML) methods (Yang et al., 2010: Yang, Lee, Goddard, & Vissher, 2011).

It is now understood that the genetic variance of complex traits is explained by small effects of many genetic variants (i.e., are highly polygenic), and that SNP markers that are common (i.e. relatively frequent) in the population should not explain a high proportion of phenotypic or genetic variance alone. The proportion of phenotypic variance explained by genome wide significant hits alone (h^2_{GWAS}) is typically very small, though this will increase as a function of increasing sample sizes. The proportion of phenotypic variance explained by all available markers (h^2_{SNP}) will be higher than from significant hits alone, though it often does not reach the theoretically upper-bound predicted by twin and family studies.

Similarly, polygenic risk scores (PGS) can be used to estimate shared genetic variance across traits by using both genome wide significant SNPs, as well as information from SNPs that

do not reach this stringent threshold. A PGS is a weighted sum of the effect of alleles on a phenotype, estimated from a GWAS. This overlap can be quantified as a genetic correlation (rG), or the overlap can be quantified as the proportion of phenotypic variance explained in a secondary trait from the effects of the genotypic markers in a discovery trait. A PGS can be used to explore genetic overlap between a trait in a discovery sample and traits (i.e., the same phenotype or across different phenotypes) in an independent, replication sample. Importantly, biases can occur when the discovery sample is small due error in SNP effect sizes and model over fitting (Wray et al., 2013), if there is sample overlap (Dudbridge, 2013, Bulik-Sulivan 2015a) or if there are significant differences across the samples such as effects of age, sex, ethnicity, or genotyping platform.

An alternative way to estimate rG across traits is to use cross-trait LD score regression (Bulik-Sullivan, 2015b). Unlike the PGS method that requires full genotype information for the secondary trait, LD score regression can be applied to datasets containing only GWAS summary statistics and is not biased by sample overlap. In standard LD score regression (Bulik-Sullivan, 2015a), the LD structure of loci (i.e. the extent to which SNPs are correlated or are inherited together) is taken to account to separate the estimate of h^2_{SNP} from bias caused by cryptic relatedness or population structure (i.e. when close relatives are recruited into the same sample or a sample has subpopulations with different allele frequencies, likely due to differences in ethnic ancestry). Such biases can inflate test statistics or estimates of traditional GWAS and GREML approaches, as well as lead to spurious associations. While these issues can be addressed using statistical controls (e.g., applying a genomic control correction or controlling for ancestry with principal components), LD score regression estimates the likelihood that SNPs are tagging causal variants (e.g., which will be dependent on the LD structure of the region) and parses these effects

from signals that are uncorrelated with LD structure. Thus, LD score regression provides a direct estimate of confounding bias and an estimate of variance explained by true polygenicity.

In this chapter, we aim to explore the genetic architecture of number of lifetime sexual partners and test the genetic overlap with behavioral disinhibition (BD) and related diseases and traits, as well as other fitness phenotypes. We use several genome wide approaches including exploring top hits from a genome wide association study in the UK Biobank and replication samples, estimating heritability explained by common SNPs estimated through LD score regression, and by exploring genetic overlap between number of lifetime partners across samples and with other related traits.

Methods

Sample

UK Biobank is a large-scale study of medical illness (e.g. with the full sample eventually totaling over 500,000) age 40-59 years. Genotypes of over 120,000 participants with thorough medical records have been released to date (Sudlow et al., 2015), though the phenotype number of lifetime sexual partners was only available for 93,625 participants and included only participants of European ancestry. Effect sizes for SNPs should be highly accurate in this well-powered sample, thus it was used as the discovery sample for GWAS and for calculating PGS predictive of number of lifetime sexual partners.

CADD/GADD included 1089 singletons (n=757 males, 332 females) ages 18-36 (*m*= 25.39, *s.d.*= 3.08) with available genotype information. The CADD sample (described previously in Chapter IV) included samples representative of state demographics and a selected sample of clinical probands selected for high BD behavior. Additional high-risk participants from the Genetics of Antisocial Drug Dependence (GADD) were also utilized. The GADD is a

collaboration between CU Boulder, CU Denver, and UC San Diego that recruited an independent replication sample for the CADD clinical participants.

Minnesota Center for Twin and Family Research (MCTFR) included 1395 singletons (n=338 males, 1057 females) ages 18-37 (*m*= 25.05, *s.d.*= 4.36). The larger sample was ascertained from Minnesota state birth records and is representative of state demographics (Miller et al. 2012), though an Enrichment Sample was added that overrepresented twins that were high on BD behavior (Keyes et al., 2009).

Measures

UK Biobank asked a single open-ended item: "About how many sexual partners have you had in your lifetime? Sexual intercourse includes vaginal, oral, or anal intercourse."

CADD/GADD scores were assessed using the quasi-continuous variable: "In your lifetime, with how many different partners have you had oral, vaginal or anal sex?" assessed on seven-point scale ("none" [0], "one" [1], "two" [2], "three-five" [3], "six-nine" [4], "ten-nineteen" [5], and "twenty or more" [6]). A subset of participants were asked the following open-ended item "In your life, with how many people have you had sex with?," for which scores were rescaled to match the existing seven point ordinal scale.

MCTFR assessment split partner items into the following four questions: 1) "How many different causal partners have you had sexual intercourse (either vaginal or anal) with in your lifetime?," 2) "How many different committed romantic partners have you had sexual intercourse (either vaginal or anal) with in your lifetime?," 3) "How many different causal partners have you oral sex with in your lifetime?," and 4) "How many different committed romantic partners have you had oral sex with in your lifetime?." It was unclear whether the same partners were reported for both penetrative sex (either vaginal or anal) and oral sex. Since

penetrative and oral sex partnerships were highly correlated (r=0.77 & r=0.78, for causal and committed partnerships respectively), we chose to use only penetrative sex as an index of risk. The final phenotype was a sum score of total number of causal and committed penetrative partners reported in lifetime. Each partnership type was assessed on seven-point scale ("none" [0], "one" [1], "two" [2], "three-five" [3], "six-nine" [4], "ten-nineteen" [5], and "twenty or more" [6]), so the resulting ordinal sum score reflected a range of partnerships rather than an actual count. This scale was identical to the scale used for CADD/GADD participants.

Genotyping

The UK Biobank Axion Array was used in the discovery sample, which captures common (minor allele frequency (MAF) >5% in European samples) and rare autosomal SNP variants (MAF<5% in European samples), multi-allelic markers, copy number variants, mitochondrial markers, and Y-chromosome markers (for more information, see Hoffman et al., 2011). However, analyses were limited to SNPs that met standard quality control (QC) thresholds. These included removing SNPs with call rates was less than 95%, minor allele frequency less than 1%, or if there was a significant deviation from Hardy–Weinberg equilibrium (P-value > 10^{-6}). Individuals were dropped if their genetic sex did not match reported gender.

Though the replication samples used different platforms (CADD/CADD: Affymetrix 6.0 platform & MCTFR: Illumina Human 660W-Quad array), both samples and SNPs were imputed through the Michigan Imputation Server resulting in over 45,000,000 variants per sample (Das et al., 2016). For more information, see McGue et al., 2013 and Derringer et al., 2015. The same QC procedures were applied to both samples.

Analyses

GWAS

Males and females were included in a single analysis in the UK Biobank sample, controlling for ancestry, batch effects, age, sex, and birth year. Test statistics were obtained for 12,191,617 common SNPs (MAF>.01).

Individual GWAS were conduced in the two replication samples and results were harmonized using Plink tools for meta-analyses (Purcell et al., 2007), which identified 6,450,299 common SNPs (MAF>.01) present in both samples. However, the total sample size of the replication samples was small (n=2,484) and was underpowered to find hits reaching genome wide significance. Rather, the replication sample was used to explore effect sizes of top SNPs in the discovery sample.

Gene and gene set analysis was completing using MAGMA (de Leeuw et al., 2015). This tool uses multiple regression to aggregate the effects of single SNP analysis at the level of a gene and gene sets that are biologically or functionally related. Top hits were verified with RegulomeDB v 1.1 (Boyle et al., 2012), a database that includes known or likely signs of DNA regulation based on public literature and databases such as the Encyclopedia of DNA Elements (ENCODE Project Consortium, 2012).

Polygenic Scores (PGSs)

PGSs were estimated from GWAS summary statistics estimated in the UK Biobank discovery sample, and applied to the independent replication samples using Plink (Purcell et al., 2007). We constructed five scores that included SNPs with p-values below the following p-value thresholds in the discovery GWAS: 0.00001, 0.01, 0.10, 0.50, and 1 (i.e., all markers), as the optimal PGS p-value threshold varies by genetic architecture of the trait (So & Sham, 2017).

Similar to methods used in previous studies (Reitveld et al., 2014), two stages of scores were estimates. The first set of scores controlled for significant phenotypic covariates (i.e. age and sex). The next set of scores controlled for confounding effects of population stratification and controlled for age, sex, and the first four principal components estimated from the genetic data.

PGSs typically do not make use of the linkage disequilibrim (LD) structure among the SNPs. Whole-genome LD score regression was used to estimate how much variance in number of lifetime partners can be explained by common SNPs (h^2_{SNP}) taking LD structure into account. LD scores were calculated using the 1000 Genomes project as a reference (1000 Genomes Project Consortium, 2012) with European samples. Due to difficulties in estimating LD in the MHC region between 26 and 33Mb on chromosome 6, these variants were removed from the analyses. Finally, we used common, well-imputed SNPS identified by the HapMap 3 reference panel to control for imputation quality (International HapMap 3 Consortium, 2010).

LD Score Regression

In addition to computing the genetic correlation between our phenotype in the discovery and replication samples, genetic correlations were computed for other BD and fitness related phenotypes. GWAS level summary statistics were accessed via LD Hub v1.4.0 (Zheng et al., 2017), a LD score regression analysis pipeline and database hosting 173 diseases and traits from databases and consortia such as the dbGAP (database of Genotypes and Phenotypes), GIANT (Genetic Investigation of ANthropometric Traits), PGC (Psychiatric Genomics Consortium), SSGAC (Social Science Genetics Association Consortium), ReproGen (Reproductive Genetics Consortium), and TAG (Tobacco and Genetics Consortium). Though an atheoretical, exploratory approach is possible with this tool, we limited analyses to phenotypes that have been linked to sexual behavior to reduce the burden of multiple testing. We included 21 diseases and or traits

including reproductive traits (i.e., age of menarche, age at menopause, age at first birth, and number of children ever born), smoking phenotypes (i.e., cigarettes smoked per day, ever vs. never smoked, current vs. former smoked, and age of smoking initiation), BD disorders and related psychopathology (i.e., ADHD and subjective well being), measures of education or cognitive ability (i.e., childhood IQ, college completion, and years of schooling), anthropometric or fitness traits (i.e., body mass index [BMI], child birth length, child birth weight, overweight, waist-to-hip ratio), and personality traits (i.e. conscientiousness, openness to experience, and neuroticism). Bonferroni corrections were applied to correct for multiple testing.

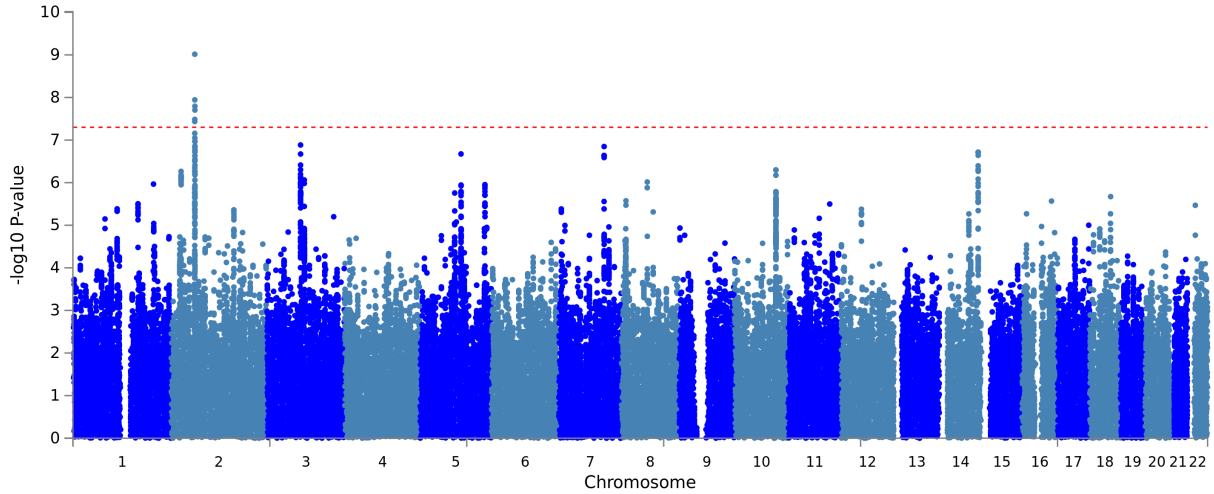
Results

GWAS

In the UK Biobank discovery sample, 7 SNPS were identified that were genome wide significant (loci with variants associated with number of lifetime partners associated at p <= 5.0e-08; See Figure 5.1). However, only two of these SNPs are independent (rs2419405, p=9.655e-10, rs4672376, p=1.603e-08; See Figure 5.2). Figures 5.1 and 5.2 were created using the functional mapping and annotation of genetic associations web-based platform (FUMA; Watanabe, Taskesen, van Bochoven, & Posthuma, 2017).

Figure 5.1

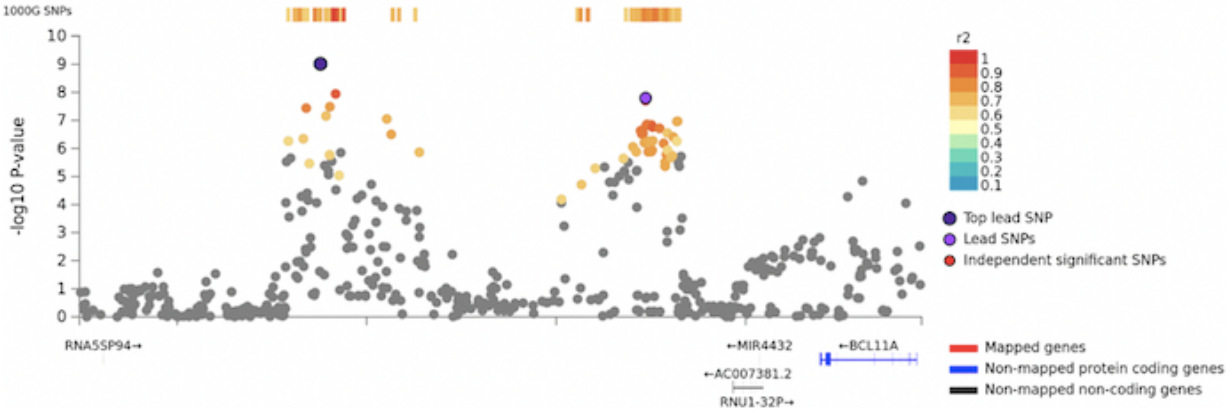
Manhattan plot for results in the UK Biobank discovery sample (n= 93,625)



Note: Points above the red line indicates SNPs that reached genome wide statistical significance (p <= 5.0e-8).

Figure 5.2

SNP plot for independent hits in UK Biobank discovery sample (n= 93,625)



Note: Top lead SNP (rs2419405) on the left, second independent SNP on the right (rs4672376).

Both lead SNPs were found in intragenic regions of chromosome 2. The top lead SNP was located near RNA5SP94, a pseudogene. The nearest gene to rs4672376 was AC007381.1, which is a long intergenic noncoding RNA (i.e., lncRNA) that is non-protein coding. The nearest

protein-coding gene to both lead SNPs was BCL11A, in a variant which has preliminary evidence for an association with sickle cell anemia (Dadheech et al., 2016).

For the top hits in the discovery, rs2419405 was not significant in the replication samples. The second lead SNP, rs4672376, reached nominal significance in the both replication samples and in the meta-analysis (i.e., not at the level of genome wide significance). However, both CADD/GADD and MCTFR were underpowered to detect genome wide significant effects (See Table 5.1).

Table 5.1

Effect sizes of top SNPS from discovery sample in replication samples

		CADD/GADD		MCTFR		META	
SNP rsID	CHR:BP	ß	p	ß	p	ß	p
rs2419405	2:60151182	-0.08	0.29	-0.07	0.34	-0.04	0.48
rs4672376	2:60494345	-0.17	0.02	0.17	0.02	-0.13	0.02

Note: Effect (ß) is scored in the direction of the effect allele in the UK Biobank. CHR = chromosome, BP= base pair.

Results from the gene-based study identified (See Table 5.2) identified seven significant genes, after accounting for the effects of all SNPs. A gene-set analysis identified two enriched pathways from the Molecular Signatures Database (MSigDB; Liberzon et al., 2011), though both did not survive correction for multiple testing (See Table 5.3). Finally, scores estimated from RegulomeDB suggested weak evidence of regulatory function.

Table 5.2

Top genes idenfitied in the discovery sample

Gene	CHR	BP Start	BP Stop	# of SNPS	Z	p
CADM2	3	84998132	86133579	303	5.24	8.136e-08
CNNM2	10	104668050	104859978	65	5.01	2.748e-07
NT5C2	10	104835940	104963056	47	4.89	4.9434e-07
AS3MT	10	104619273	104671656	23	4.89	4.9673e-07
DAGLB	7	6438757	6533821	52	4.83	6.8471e-07
C10orf32-ASMT	10	104604029	104671656	31	4.79	8.1819e-07
MSRA	8	9901778	10296401	329	4.59	2.2363e-06

Note: CHR = chromosome, BP= base pair.

Table 5.3

Top gene sets in the discovery sample

Gene Set	# of Genes	$ß_{standardized.}$	p	$p_{corrected}$
Fontaine_thyroid_tumor_uncertain_ malignancy_up	32	0.566	6.184e-05	0.67
GO_Regulation_of_Neuron _ Differentiation	528	0.141	8.726e-05	0.95

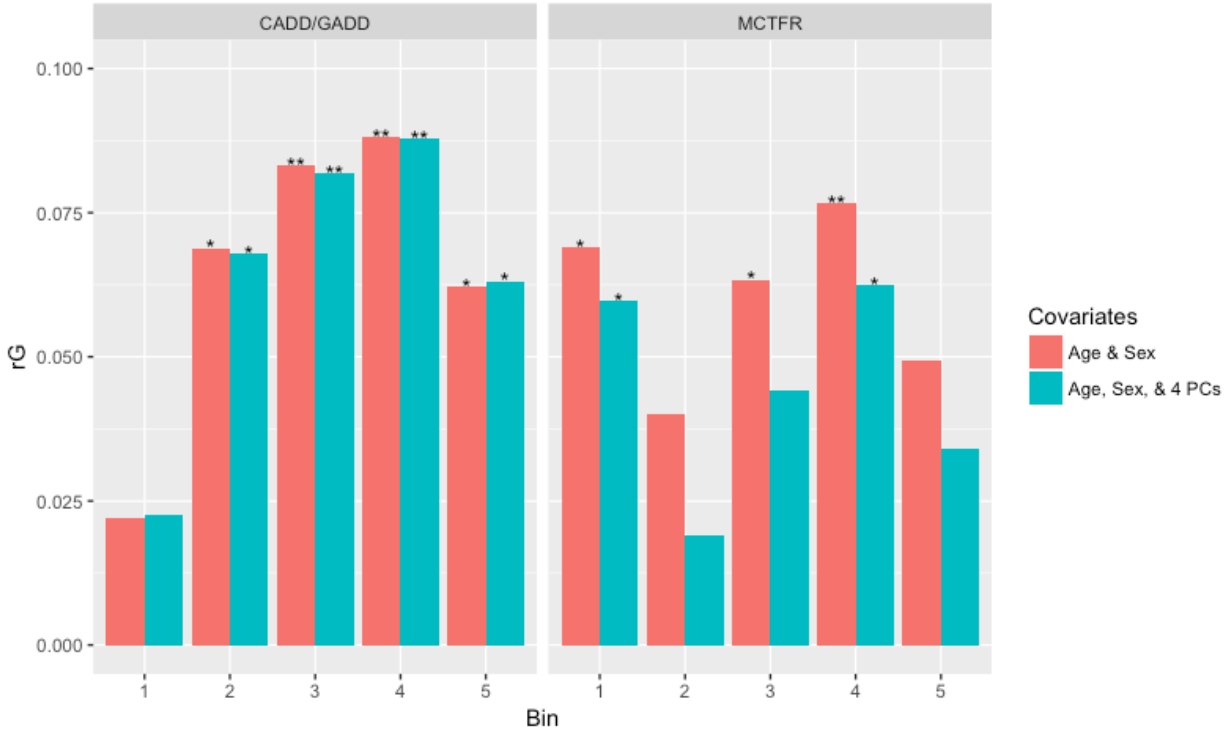
Polygenic Scores (PGSs)

Each PGS was correlated with the phenotype in the discovery sample (i.e., rG). All estimated rGs were in the positive direction (i.e. the SNPs that were associated with increased number of lifetime partners in the UK Biobank sample were on average predictive of increased number of lifetime partners in the replication samples).

The first stage of scores controlled for age and sex, which were significant predictors of number of lifetime partners in both samples. Of the five p-value threshold bins, four were significant at $p<.05$ in the CADD/GADD sample and three were significant in the MCTFR (See Figure 5.3). After controlling for population structure, four bins remained significant in the CADD/GADD sample while only three were significant at $p<.05$ in MCTFR. In the more stringent stage, the most predictive bins in both samples included SNPs with p-values <0.50. While these correlations were significant, the scores explained only a small amount of variance in the phenotypes of the respective replication samples ($R^2= 0.0077$ in CADD/GADD and $R^2= 0.0039$ in MCTFR).

Figure 5.3

Genetic correlations for number of lifetime partners in replication samples



Note: P-value thresholds for Bins 1-5 were as follows: <0.00001, <0.01, <0.10, <0.50, & <1.00. rG significance denoted as p<.05 (*) and p<.01(**).

LD Score Regression

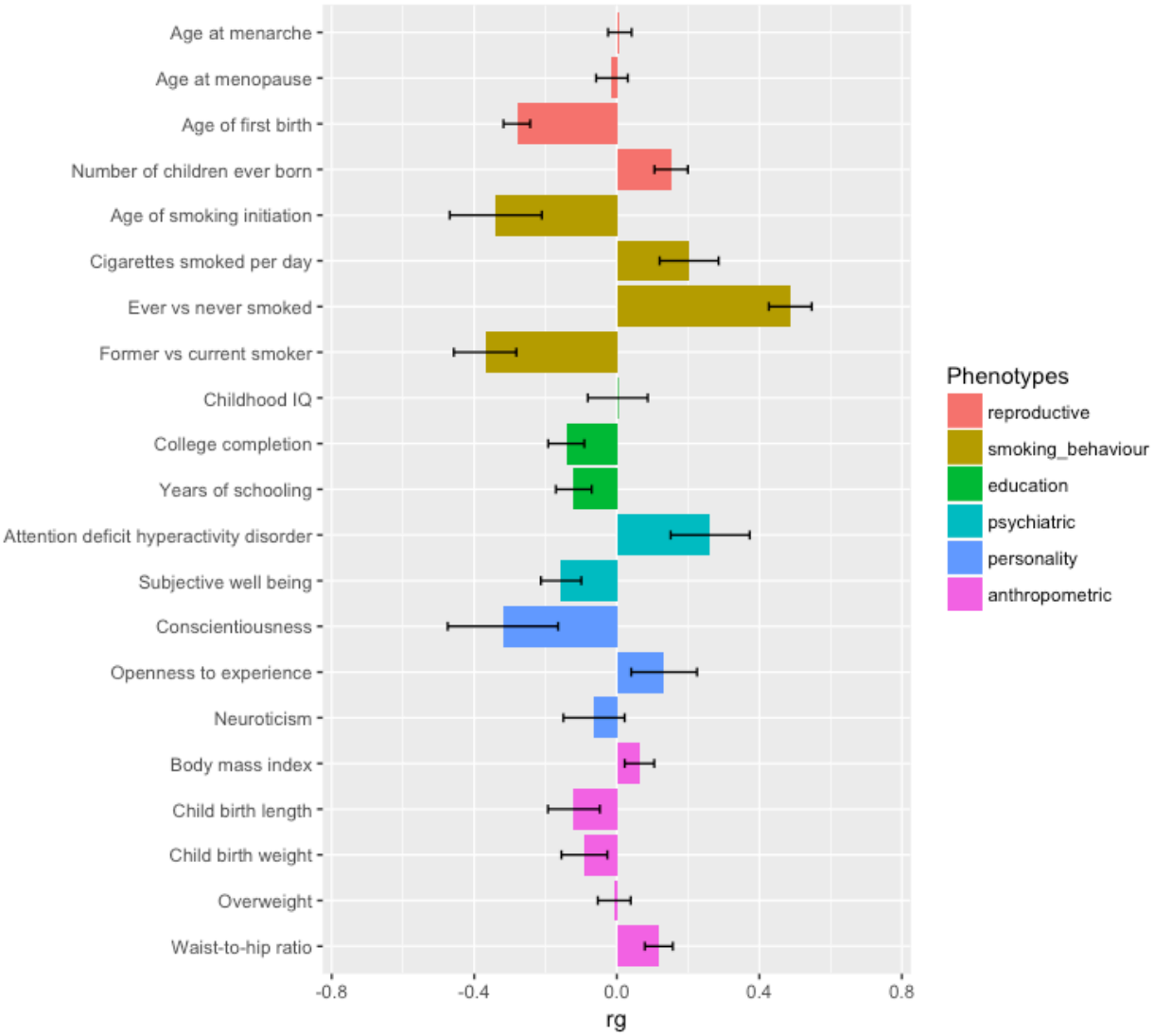
After merging with the reference panel, the number of SNPs that remained in the analyses was 1,183,388 for UK Biobank, 1,039,982 for CADD/GADD, and 994,958 for MCTFR. In the UK Biobank, variation tagged by common SNPs explained approximately 10% of the overall phenotypic variance in number of lifetime sexual partners (h^2_{SNP}=0.1023, *s.e.*= 0.0075). Caution should be taken when interpreting estimates in the replication samples, given the low sample sizes. Heritability was estimated in the CADD/GADD sample (h^2_{SNP}=.2796, *s.e.*= 0.3968); however, the estimate in the MCTFR was negative (h^2_{SNP}= -0.1806, *s.e.*= 0.2957) which is suggestive of sampling error. Thus, the genetic correlation between number of lifetime

partners across samples could only be estimated in the CADD/GADD sample (rG= 0.604, s.e.= 0.4649, p = 0.1939).

In addition to estimating the genetic correlation with number of lifetime partners in the replication samples, genetic correlations were estimated for BD related diseases and traits (e.g. smoking, psychiatric, personality), and other fitness phenotypes (e.g., reproductive, educational, or anthropomorphic). All phenotypes were from independent samples, with the exception of age of first birth and number of children ever born. Across these 21 measures, we identified twelve significant genetic correlation at $p<.05$ (See Figure 5.4). The following Bonferonni corrections for multiple testing were applied: 1) one threshold accounting for the number of tests conducted ($p = 2.34e{-3}$), and 2) a second more stringent threshold accounting for the number of potential tests possible through LD Hub ($p = 2.89e{-4}$). Three correlations survived both thresholds (i.e., ever vs. never smoked [rG= 0.49], age at first birth [rG=-0.28], and former vs. current smoker [RG=-0.37]). The pattern of significant correlations, based on SNP-level analyses, is consistent with the hypothesis that RSB and BD share common genetic influences.

Figure 5.4

Genetic correlations with estimated with LD score regression



Note: Error bars indicate standard error (*s.e.*).

Discussion

The goal of the study was to use genome wide approaches to examine the genetic architecture of number of lifetime sexual partners, a measure of risky sexual behavior that is moderately to highly heritable and associated with negative health outcomes.

Results from the GWAS in the UK Biobank identified two independent SNPs and several potential genes associated with number of lifetime partner, though the biological and functional

importance of the identified region remains unknown. Interestingly, the CADM2 gene identified in the gene based analyses was recently associated with number of children and self-reported risk taking propensity, though these results were found in the UK Biobank sample (Day et al., 2016). Future replication studies, and reanalysis with the upcoming release of data from 500,000 participants in the UK Biobank, may identify additional loci and genes of importance.

The results also suggest that number of lifetime sexual partners is highly polygenic. Further pathway analyses will help to aggregate the effects of SNPs across the genome. Methods such as polygenic scores (PGSs) and LD score regression will be required to test shared genetic etiology with related traits.

We calculated several PGSs to test how well the genetic markers predicting number of lifetime sexual partners in the discovery sample predicted this phenotype in our replication samples. The first stage of PGS calculation (performed in Plink; Purcell et al, 2007) included age and sex and important phenotypic covariates. The second stage regression analysis included 4 PCs to control for population stratification (e.g., genetic ancestry). For each set of scores, several bins of p-value thresholds were utilized. Including SNPs with p-values < 0.50 provided the most predictive PGS in both samples, suggesting these scores had the best signal-noise ratio. Though the estimates of rG were significant, each PGS explained less than 1% of the phenotypic variance in lifetime number of sex partners in the replication samples.

Other studies have found similarly low PGS prediction as well, and these effect sizes may be expected for the following reasons. The estimate of h^2_{snp} from LD score regression within the UK Biobank discovery sample analysis was small at approximately 10%. Though heritability estimates from LD score regression may be slightly lower than other GREML methods (Lee, Vattikuti, & Chow, 2016), this estimate is far lower than what is predicted by twin and family

studies. Thus, there is a low upper bound for prediction across samples (i.e. if 10 % phenotypic variance in number of lifetime sexual partners could be explained by variance in common SNPs in the discovery sample, these SNPs should predict 10% or less of the variance in an independent sample; Dudbridge, 2013; Chatterjee et al., 2013) Additionally, there are several important differences across our samples that might limit generalizability including cultural differences, age effects, birth cohort effects, or differences in ancestry (Martin et al., 2017). Finally, the PGS estimates are based only on genetic variance tagged by common SNPs excluding effects of rare variants, structural variants (such as copy number variants), variants on the sex chromosomes, or effects of gene x environment interactions and gene x gene interactions (i.e. epistasis).

LD score regression was used as a second method to explore genetic overlap. Indeed, there were positive rG between number of lifetime sexual partners in the discovery sample and the CADD/GADD sample. Using publically available summary statistics from BD and fitness phenotypes from related GWAS, we found that genetic loci predicting number of lifetime sexual partners were also significant predictors of two smoking phenotypes and one reproductive trait after correcting for multiple testing. The genetic association with smoking phenotypes was expected given the correlation between number of lifetime partners and BD. Though the rG with age of smoking initiation and number of cigarettes per day did not survive multiple test correction, Bulik et al. has previously reported only weak, non-significant genetic correlations between age of smoking initiation and the other smoking phenotypes in the TAG consortium (2015b). Substantial shared overlap was also expected between number of lifetime partners and reproductive traits. Though age of sexual initiation was not available via LD Hub, a recent study using UK Biobank data estimated a significant negative genetic correlation between age of sexual initiation and number of lifetime sexual partners (Day et al., 2006). Overall, these results

are in line with twin and family research that suggest RSB has common genetic antecedents with BD.

In summary, the purpose of this study was to explore the genetic architecture of number of lifetime partners and the genetic overlap with related traits. Though some significant loci were identified, larger sample sizes and replication studies will be needed. As a future direction, these analyses will be redone with the release of 500,000 genotypes in the UK Biobank sample. Finally, additional or larger independent samples will be added for the sake of replication.

CHAPTER VI

OVERALL DISCUSSION

A heritable common factor, often referred to as behavioral disinhibition (BD), explains comorbidity of many impulsive behaviors including substance use disorders, antisocial, and novelty seeking behavior. However, there is no consensus on whether risky sexual behavior (i.e., those behaviors leading to negative health outcomes) should be considered a central indicator of this factor. Various measures of risky sexual behavior are sometimes used as components of BD, and many RSB measures are highly correlated with measures of BD that exclude RSB indicators. Given the high heritability of BD, RSB should be correlated with BD mainly due to common genetic influences if it is truly is another component of BD.

The overall goal of this dissertation was demonstrate the shared etiology (i.e. genetic and environmental overlap) of risky sexual behavior with other indicators of behavioral disinhibition using a variety of developmental and multivariate biometric models, as well as quantitative genome wide methods.

Chapter II

This study used developmental biometric modeling and an adoption design to test trends in substance use behavior (a central component of BD). Primary contributions to the literature included corroborating existing cross sectional twin findings with 1) adoption results, which provide a direct estimate of shared environmental influences, and 2) dense longitudinal measures of substance use. Univariate biometric estimates across adolescence (i.e., ages 14-18) indicated moderate hertiabilites for liability to use cigarettes, alcohol and marijuana. Heritability decreased in magnitude for cigarette and alcohol use across adolescence, but increased for marijuana use. Shared environmental influences were relatively modest for cigarette use/no use across

adolescence. For alcohol use, there was a trend for increasing shared environmental influences with the greatest influence at age 18. The adoptive and control sibling correlations for quantity/frequency of substance use generally suggested genetic influences, with only modest effects of the shared environment, particularly at early ages when prevalence of use was lower.

Developmental trends were tested with a series of independent pathway models. Genetic and shared environmental influences on all substances were mostly common across adolescence (i.e., all age-specific influences could be dropped from the models). Additionally, the magnitude of these common influences were fairly stable across time (i.e., path loadings could be constrained to be equal across age). However, significant age-specific non-shared environmental influences were identified, which could suggest measurement error that is not correlated across age. Non-shared environmental influences that were common across age were highly variable for all substances. These results suggest that influences that make family members more similar are stable through adolescence; however, it is possible that we were unable to detect trends with small effects due to limited lower. Future studies could increase power with the addition of other types of genetically related pairs (i.e., twins or other extended family members).

While this study did not explore RSB directly, these biometric models could be applied to explore the developmental trends of adolescent sexual behavior in order to corroborate twin findings with an adoption design using dense measures of behavior across adolescence.

Chapter III

This study was the first to use multivariate biometric models in twins to test whether sex under the influence of alcohol or drugs had a direct causal effect on lifetime number of sexual partners (a proxy for RSB), improving upon the limitations of cross sectional or experimental research. Previous studies have typically been unable to control for important third variables that

may influence both use of substances and sexual behavior. For instance, we expected that influences on BD traits may independently influence both sex under the influence and lifetime number of partners. As such, we found evidence that higher scores on our composite measure of sex under the influence (e.g. how frequently substance use influenced sexual decision making) did predict higher number of sexual partners; however, increases in drug and alcohol use, more generally, also predicted higher number of sexual partners.

We explored this relationship using biometric models (which decompose covariance between sex under the influence and number of lifetime partners into additive genetic, shared environmental, and nonshared environmental factors) and direction of causation models in twins. For males and females, covariance between sex under the influence and number of lifetime partners was explained by both additive genetic and nonshared environmental factors. As this genetic overlap could reflect true pleiotropy or mediated pleiotropy (e.g., when one trait is causally related to a second trait so that the genetic factors for the first trait are indirectly associated with the second), this pattern alone was not informative about causality.

A trivariate Cholesky decomposition model was used to explore the nature of the covariance between sex under the influence and number of lifetime sexual partners, when controlling for general substance use. After controlling for substance use in general, the genetic covariance specific to sex under the influence and number of lifetime partners was reduced in males but remained significant. That is, there were some genetic factors that influence substance use (within and outside of sexual contexts) and number of lifetime sexual partners, as well as some genetic influences specific to sex under the influence and lifetime number of partners. In females, much of the relationship between sex under the influence and number of lifetime partners was mediated through influences in common with drug use in general.

Finally, results from the direction of causation models suggested that sex under the influence did not have a direct causal effect on lifetime number of sexual partners; rather the reverse pattern better fit the data. However, this result does not prove reverse causation (i.e., these results are necessary but not sufficient for causality), as unmeasured third variables or biases due to item measurement could account for this finding. These results should be interpreted as conflicting support for a casual model (i.e., it is unlikely that substance use during sex is a primary *cause* of risky sexual behavior).

It is important to note that our measure of sex under the influence differed from some other research. Rather than measuring the frequency of non-sober sex, our measure may be capturing the frequency in which substance use impaired sexual decision making (presumably, leading to increased risk). Additionally , the measure may assess a tendency to use substances in order to facilitate sexual encounters, or it may capture a tendency to endorse sex under the influence measures as a form of post-hoc rationalization to remove dissonance or social shame of RSB. Motivations for endorsing this item are explored further in Chapter IV.

Chapter IV

The purpose of this chapter was threefold. The first goal was to explore survey responses to our measure of sexual behavior (i.e., the M-RBQ) in parallel with a literature review of RSB, in order to create a revised version of the instrument that was efficient and informative. The second purpose was to inform and clarify the meaning and interpretation of the variables selected for analysis in Chapter III (i.e. number of lifetime sexual partners and sex under the influence). Several lines of evidence suggested that number of lifetime partners is a good proxy measure of RSB. Additionally, it was clear from a pilot study of undergraduates that motivations for endorsing sex under the influence of drugs and alcohol were complex. There was some evidence

that use of drugs and alcohol during sexual decision making did not necessarily lead to *riskier* sexual behavior.

Finally, estimates from GWAS, and genome wide approaches more generally, can be limited by poorly measured phenotypes, phenotypes that are highly heterogeneous, or phenotypes that have low heritability (McCarthy & Hirschhorn, 2008; Manchia et al., 2013). Thus, this chapter aimed to identify an optimal phenotype for use in Chapter V. Our selection criteria for an optimal phenotype included 1) moderate to high heritability, 2) reliable with low measurement error [i.e. easily standardized across samples], 3) little heterogeneity or sex-limitation, and 4) a continuous and normal distribution. Additionally, we expected that any measure of RSB should also be predictive of other risk measures, BD, and negative health outcomes (in accordance with our definition of RSB). Given these criteria, number of lifetime sexual partners was selected for use in the final study, which used genome wide approaches.

Chapter V

Genome wide methods were used to explore the genetic architecture of number of lifetime sexual partners and test the genetic overlap with behavioral disinhibition (BD) and related diseases and traits, as well as other fitness phenotypes. Two independent SNPs were associated with number of lifetime partners in the UK Biobank, though the biological and functional significance of these hits were not clear. As expected, this complex phenotype was highly polygenic. Phenotypic variance explained by genetic variation tagged by common SNPs (i.e., h^2_{SNP}) was estimated using LD score regression. Within the UK Biobank discovery sample approximately 10% of the variance in number of lifetime sexual partners was explained, which is significantly lower than estimates from twin and family studies.

Several polygenic scores (PGSs) were estimated using five p-value threshold bins. In two independent samples, the score with the most predictive power explained small but significant phenotypic variance. Finally, significant genetic correlations between number of lifetime sexual partners with smoking and reproductive phenotypes were identified. These findings provide the first SNP-based evidence in support of the hypothesis that RSB and BD share a common genetic etiology.

The upcoming release of the UK Biobank (n=500,000) will be used in a future analysis, to improve the overall power of the discovery GWAS. This should increase the predictive power of PGSs, as well as create more accurate estimates of h^2_{SNP} and rG using LD score regression. While this investigated the overlap with a number of behaviors associated with BD, we were unable to examine the genetic overlap with a wide range of BD indicators. The following release will include additional measures of risk and BD traits, for which the genetic overlap with number of lifetime sexual partners is unknown.

Summary

Together these studies have explored the shared etiology of risky sexual behavior and behavioral disinhibition. This work suggests number of lifetime measures may be an optimal way to assess RSB in young adult samples and that this variable has genetic influences common to the heritable factor of BD. Future work is needed to determine the underlying biological and environmental mechanisms contributing to this overlap.

CHAPTER VII

REFERENCES

1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56-65.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: a review and analysis of empirical efforts. *Psychological Bulletin*, *85*, 1275-1301.

Agrawal, A., & Lynskey (2006). The genetic epidemiology of cannabis use, abuse, and dependence. *Addiction, 101,* 801-812.

Agrawal, A., Few, L., Nelson, E. C., Deutsch, A., Grant, J. D., Bucholz, K. K., Madden, P. A. F., Health, A., & Lynskey, M. T. (2016). Adolescent cannabis use and repeated voluntary unprotected sex in women. *Addiction*, *111,* 2012-2020.

The Alan Guttamacher Institute (2016). American teen's sexual and reproductive health. Retrieved from https://www. guttmacher.org/fact-sheet/american-teens-sexual-andreproductive-health

Allen, N. E., Sudlow, C., Peakman, T., (2014). UK Biobank data: Come and get it. *Science Translational Medicine, 6,* 224ed4.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Aral, S. O. (2001). Sexually transmitted diseases: magnitude, determinants and consequences. *International Journal of STD & AIDS*, *12*(4), 211-215.

Bailey, S. L., Pollock, N. K., Martin, C. S., & Lynch, K. G. (1999). Risky sexual behaviors among adolescents with alcohol use disorders. *Journal of Adolescent Health*, *25*, 179-181.

Baker, J. H., Maes, H. H., Larsson, H., Lichtenstien, P., & Kendler, K. S. (2011). Sex difference and developmental stability in genetic and environmental influences on psychoactive substance consumption from early adolescence to young adulthood. *Psychological Medicine, 9,* 1907-1916.

Barban, N., Jansen, R., de Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., et al. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics*, *48,* 1462-1472.

Basen-Engquist, K., Mâsse, L. C., Coyle, K., Kirby, D., Parcel, G., Banspach, S., & Nodora, J. (1998). Sexual risk behavior beliefs and self-efficacy scales. *Handbook of Sexuality-Related Measures*, 541-544.

Basen-Engquist, K., Masse, L. C., Coyle, K., Kirby, D., Parcel, G. S., Banspach, S., & Nodora, J. (1999). Validity of scales measuring the psychosocial determinants of HIV/STD-related risk behavior in adolescents. *Health Education Research*, *14*, 25-38.

Beadnell, B., Morrison, D. M., Wilsdon, A., Wells, E. A., Murowchick, E., Hoppe, M., Gillmore, M. R., & Nahom, D. (2005). Condom use, frequency of sex, and number of partners: Multidimensional characterization of adolescent sexual risk‐taking. *Journal of Sex Research*, *42*, 192-202.

Beck, K.H., Thombs, D.L., Mahoney, C.A., & Fingar, K.M. (1995). Social Context and sensation seeking: gender differences in college students drinking motivations. *Substance Use & Misuse, 30,* 1101-1115.

Bennett, S. N., Caporaso, N., Fitzpatrick, A. L., Agrawal, A., Barnes, K., Boyd, H. A., Cornelis, M.C., Hansel, N.N., Heiss, G., Heit, J.A., Kang, J. H., Kittner, S.J., Kraft, P., Lowe, W., Marazita, M. L., Monroe, K. R., Pasquale, L. R., Ramos, E. M., van Dam, R. M., Urden, J., Williams, K., & the GENEVA Consortium (2011). Phenotype harmonization and cross‐study collaboration in GWAS consortia: the GENEVA experience. *Genetic Epidemiology*, *35*, 159-173.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in the heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics, 10,* 423-433.

Bernreuter, R. G. (1933). Validity of the personality inventory. *Personnel Journal*, *11,* 383-386.

Boardman, J. D., & Fletcher, J. M. (2015). To cause or not to cause? That is the question, but identical twins might not have all of the answers. *Social Science & Medicine*, *127*, 198-200.

Boislard, M. A., van de Bongardt, D., & Blais, M. (2016). Sexuality (and lack thereof) in adolescence and early adulthood: A review of the literature. *Behavioral Sciences*, *6*, 8; doi:10.3390/bs6010008.

Boomsma, D. I., & Molenaar, P. C. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behavior Genetics*, *17*, 111-123.

Booth R.E., Corsi, K.F., & Mikulich-Gilbertson, S.K. (2004). Factors associated with methadone maintenance treatment retention among street-recruited injection drug users. *Drug and Alcohol Dependence, 74*,177-185.

Bornovalova, M. A., Hicks B. M., Iacono, W. G., & McGue, M. (2010). Familial transmission and heritability of childhood disruptive disorders. *American Journal of Psychiatry, 167,* 1066–1074.

Brener, N. D., Kann, L., Shanklin, S., Kinchen, S., Eaton, D. K., Hawkins, J., & Flint, K. H. (2013). Methodology of the youth risk behavior surveillance system—2013. *MMWR Recommendations and Reports*, *62*, 1-20.

Bricker, J. B., Stallings, M. C., Corley, R. P., Wadsworth, S. J., Bryan, A., Timberlake, D. S., Hewitt, J.K., A. Caspi, S. M., Hofer, S. M., Rhea, S. A., & DeFries, J. C. (2006). Genetic and environmental influences on age at sexual initiation in the Colorado Adoption Project. *Behavior Genetics*, *36*, 820-832.

Brown, J. L., Sales, J. M., DiClemente, R. J., Salazar, L. F., Vanable, P. A., Carey, M. P., Brown, L. K., Romer, D., Valois, R. F., & Stanton, B. (2012). Predicting discordance between self-reports of sexual behavior and incident sexually transmitted infections with African American female adolescents: results from a 4-city study. *AIDS and Behavior*, *16*, 1491-1500.

Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research*, *36*, 292-297.

Buchanan, J. P., McGue, M., Keyes, M., & Iacono, W. G. (2009). Are there shared environmental influences on adolescent behavior? Evidence from a study of adoptive siblings. *Behavior Genetics, 39,* 532-540.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., et al. (2015b). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47, 1236-1241.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A.L, & Neale, B.M. (2015a). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*, 291-295.

Caminis, A., Henrich, C., Ruchkin, V., Schwab-Stone, M., & Martin, A. (2007). Psychosocial predictors of sexual initiation and high-risk sexual behaviors in early adolescence. *Child and Adolescent Psychiatry and Mental Health*, doi: 10.1186/1753-2000-1-14

Catania, J. A. (1999). A framework for conceptualizing reporting bias and its antecedents in interviews assessing human sexuality. *Journal of Sex Research*, *36*, 25-38.

Cassell, J. A., Mercer, C. H., Imrie, J., Copas, A. J., & Johnson, A. M. (2006). Who uses condoms with whom? Evidence from national probability sample surveys. *Sexually Transmitted Infections*, *82*, 467-473.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, *60*, 345-375.

Catania, J. A., Coates, T. J., Stall, R., Turner, H., Peterson, J., Hearst, N., Dolcini, M. M., Hudes, E., Gagnon, J., Wiley, J., & Groves, R. (1992). Prevalence of AIDS-related risk factors and condom use in the United States. *Science, 258,* 1101-1106.

Catania, J. A., Gibson, D. R., Chitwood, D. D., & Coates, T. J. (1990). Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin, 108,* 339–362.

Cavazos-Rehg, P. A., Spitznagel, E. L., Bucholz, K. K., Nurnberger, J., Edenberg, H. J., Kramer, J. R., Kuperman, S., Hasselbrock, V., & Bierut, L. J. (2010). Predictors of sexual debut at age 16 or younger. *Archives of Sexual Behavior*, *39*, 664-673.

Centers for Disease Control and Prevention (2011). 10 ways STDs impact women differently from men. Retrieved from https://www.cdc.gov/std/health-disparities/stds-women-042011.pdf

Centers for Disease Control and Prevention (2017). Anal sex and HIV risk. Retrieved from https://www.cdc.gov/hiv/risk/analsex.html

Centers for Disease Control and Prevention (2013). Sexual activity between males and females: Vaginal sexual intercourse. Key Statistics from the National Survey of Family Growth. Retrieved from http://www.cdc. gov/nchs/nsfg/key_statistics/s.htm#vaginalsexual

Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2015. Atlanta: U.S. Department of Health and Human Services; 2016.

Cha, S., Masho, S. W., & Mezuk, B. (2016). Age of sexual debut and cannabis use in the United States. *Substance Use & Misuse*, *51*, 439-448.

Charnigo, R., Noar, S. M., Garnett, C., Crosby, R., Palmgreen, P., & Zimmerman, R. S. (2013). Sensation seeking and impulsivity: combined associations with risky sexual behavior in a large sample of young adults. *Journal of Sex Research*, *50*, 480-488.

Chandra, A., Copen, C. E., & Mosher, W. D., Copen, C., Division of Vital Statistics, National Center for Health Statistics, & Sionean, C. (2011). Sexual behavior, sexual attraction, and sexual identity in the United States: Data from the 2006–2008 National Survey of Family Growth. *National Health Statistics Report, 36, 1-36.*

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J. H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, *45*, 400-405.

Cherkas, L. F., Oelsner, E. C., Mak, Y. T., Valdes, A., & Spector, T. D. (2004). Genetic influences on female infidelity and number of sexual partners in humans: a linkage and association study of the role of the vasopressin receptor gene (AVPR1A). *Twin Research and Human Genetics*, *7*, 649-658.

Choi, K. H., Catania, J. A., & Dolcini, M. M. (1994). Extramarital sex and HIV risk behavior among US adults: results from the National AIDS Behavioral Survey. *American Journal of Public Health*, *84*, 2003-2007.

Cochran, S. D., & Mays, V. M. (1990). Sex, lies, and HIV. *New England Journal of Medicine*, *322*, 774-775.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P.K., & Seigal, M.K. (2013). Heritability and the equal environment assumption: evidence from multiple samples of misclassified twins. *Behavior Genetics, 43,* 415-426.

Copen, C. E., Chandra, A., & Febo-Vazquez, I. (2016). Sexual Behavior, Sexual Attraction, and Sexual Orientation Among Adults Aged 18-44 in the United States: Data From the 2011-2013 National Survey of Family Growth. *National Health Statistics Reports*, (88), 1-14.

Corbin, W., & Fromme, K. (2002). Alcohol use and serial monogamy as risks for sexually transmitted diseases in young adults. *Health Psychology, 21,* 229-236.

Cooper, M. L., Wood, P. K., Orcutt, H. K., & Albino, A. (2003). Personality and the predisposition to engage in risky or problem behaviors during adolescence. *Journal of Personality and Social Psychology*, *84*, 390-410.

Crowley, T. J., Mikulich, S. K., Ehlers, K. M., Hall, S. K., & Whitmore, E. A. (2003). Discriminative validity and clinical utility of an abuse-neglect interview for adolescents with conduct and substance use problems. *American Journal of Psychiatry*, *160*, 1461-1469.

Croyle, R., & Loftus, E. (1993). Recollection in the kingdom of AIDS. In D. Ostrow & R. Kessler (Eds.), Methodological issues in AIDS behavioral research (pp. 163-180). New York: Plenum. Fortenberry.

Dadheech, S., Madhulatha, D., Jain, S., Joseph, J., Jyothy, A., & Munshi, A. (2016). Association of BCL11A genetic variant (rs11886868) with severity in β-thalassaemia major & sickle cell anaemia. *The Indian Journal of Medical Research*, *143*, 449-454.

Das, S., Forer L, Schönherr, S., Sidore ,C., Locke, A. E., Kwong, A., Vrieze, S., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics, 48,* 1284–1287.

Day, F. R., Helgason, H., Chasman, D. I., Rose, L. M., Loh, P. R., Scott, R. A., Helgason, A., Kong, A., Masson, G., Magnusson, O.T., Gudbjartsson, D., Thorsteinsdottir, U., Burning, J. E., Ridker, P. M., Sulem, P., Sefansson, K., Ong, K. K., Perry, J. R, B. (2016) Physical and neurobehavioral determinants of reproductive onset and success. *Nature Genetics, 48,* 617-623.

de Leeuw C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, *11*, e1004219.

Dir, A. L., Coskunpinar, A., & Cyders, M. A. (2014). A meta-analytic review of the relationship between adolescent risky sexual behavior and impulsivity across gender, age, and race. *Clinical Psychology Review*, *34*, 551-562.

Distel, M. A., Vink, J. M., Bartels, M., Van Beijsterveldt, C. E. M., Neale, M. C., & Boomsma, D. I. (2011). Age moderates non-genetic influences on the initiation of cannabis use: a twin-sibling study in Dutch adolescents and young adults. *Addiction*, *106*, 1658–66.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research*, *7*, 659-669.

Derringer, J., Corley, R. P., Haberstick, B. C., Young, S. E., Demmitt, B. A., Howrigan, D. P. Kirkpatrick, R.M, Iacono, W.G., McGue, M., Keller, M.C., Brown, S., Tapert, S., Hopfer, C. J, Stallings, M.C., Crowley, T.J., Rhee, S.H., Krauter, K., Hewitt, J.K., & McQueen, M B., (2015). Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behavior Genetics*, *45*, 375-381.

Dermen, K.H., & Cooper, L.M. (2000). Inhibition conflict and alcohol expectancy as moderators of alcohol's relationship to condom use. *Experimental and Clinical Psychopharmacology, 8,* 198-206.

Desiderato, L. L., & Crawford, H. J. (1995). Risky sexual behavior in college students: Relationships between number of sexual partners, disclosure of previous risky behavior, and alcohol use. *Journal of Youth and Adolescence*, *24*, 55-68.

Deutsch, A. R., Slutske, W. S., Heath, A. C., Madden, P. A. F., & Martin, N. G. (2014). Substance use and sexual intercourse onsets in adolescence: A genetically informative discordant twin design. *The Journal of Adolescent Health*, *54*, 114-116.

Donahue, K. L., D'Onofrio, B. M., Lichtenstein, P., & Långström, N. (2013). Testing putative causal associations of risk factors for early intercourse in the Study of Twin Adults: Genes and Environment (STAGE). *Archives of Sexual Behavior*, *42*, 35-44.

Donahue, K. L., Lichtenstein, P., Långström, N., & D'Onofrio, B. M. (2013). Why does early sexual intercourse predict subsequent maladjustment? Exploring potential familial confounds. *Health Psychology*, *32*, 180-189.

Donohew, L., Zimmerman, R., Cupp, P. S., Novak, S., Colon, S., & Abell, R. (2000). Sensation seeking, impulsive decision-making, and risky sex: Implications for risk-taking and design of interventions. *Personality and Individual Differences*, *28*, 1079-1091.

Drumright, L. N., Gorbach, P. M., & Holmes, K. K. (2004). Do people really know their sex partners?: Concurrency, knowledge of partner behavior, and sexually transmitted infections within partnerships. *Sexually Transmitted Diseases*, *31*, 437-442.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genetics, 9,* e1003348.

Dunne, M. P., Martin, N. G., Statham, D. J., Slutske, W. S., Dinwiddie, S. H., Bucholz, K. K., Madden, P. A. F., & Heath, A. C. (1997). Genetic and environmental contributions to variance in age at first sexual intercourse. *Psychological Science*, *8*, 211-216.

Dunne, E. F., Unger, E. R., Sternberg, M., McQuillan, G., Swan, D. C., Patel, S. S., & Markowitz, L. E. (2007). Prevalence of HPV infection among females in the United States. *JAMA*, *297*, 813-819.

Eaton, N. R., Thompson Jr, R. G., Hu, M. C., Goldstein, R. B., Saha, T. D., & Hasin, D. S. (2015). Regularly drinking alcohol before sexual activity in a nationally representative sample: prevalence, sociodemographics, and associations with psychiatric and substance use disorders. *American Journal of Public Health*, *105*, 1387-1393.

Edwards, J. M., Iritani, B. J., & Hallfors, D. D. (2006). Prevalence and correlates of exchanging sex for drugs or money among adolescents in the United States. *Sexually Transmitted Infections*, *82*, 354-358.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57-74.

Epstein, M., Bailey, J.A., Manhart, L.E., Hill, K.G., & Hawkins, J.D. (2014) Sexual risk behavior in early adulthood: Broadening the scope beyond early sexual initiation. *The Journal of Sex Research*, DOI: 10.1080/00224499.2013.849652

Epstein, M., Bailey, J.A., Manhart, L.E., Hill, K.G., Hawkins, J.D., Haggerty, K.P., Catalano, R.F., (2013). Understanding the link between early sexual initiation and later sexually transmitted infection: Test and replication of two longitudinal studies. *Journal of Adolescent Health,* doi:10.1016/j.jadohealth.2013.09.016

Erens, B., Phelps, A., Clifton, S., Mercer, C. H., Tanton, C., Hussey, D., Sonnenberg, P., Mcdowall, W., Field, N., Datta, J., Mitchell, K., Copas, A. J., Wellings, K., & Johnson, A.M. (2013). Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Sexually Transmitted Infections*, 2, 84-89.

Evans, L., Tahmasbi, R., Vrieze, S., Abecasis, G., Das, S., Bjelland, D., deCandia, T., Haplotype Reference Consortium, Goddard, M., Neale, B., Yang, J., Visscher, & Keller, M. (2017). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *bioRxiv*, 115527.

Fenton, K. A., Johnson, A. M., McManus, S., & Erens, B. (2001). Measuring sexual behaviour: methodological challenges in survey research. *Sexually Transmitted Infections*, *77*, 84-92.

Fernández, M. I., Perrino, T., Bowen, G. S., Royal, S., & Varga, L. (2003). Repeat HIV testing among Hispanic men who have sex with men—a sign of risk, prevention, or reassurance?. *AIDS Education and Prevention*, *15*, 105-116.

Finer, L. B., & Zolna, M. R. (2011). Unintended pregnancy in the United States: incidence and disparities, 2006. *Contraception*, *84*, 478-485.

Finer, L. B., & Zolna, M. R. (2014). Shifts in intended and unintended pregnancies in the United States, 2001–2008. *American Journal of Public Health*, *104*, S43-S48.

Fisher, T. D., Davis, C. M., Yarber, W. L., & Davis, S. L. (2013). *Handbook of sexuality-related measures*. Routledge.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18,* 37-68.

Freimer, N., & Sabatti, C. (2003). The human phenome project. *Nature Genetics*, *34*, 15-21.

Forhan, S. E., Gottlieb, S. L., Sternberg, M. R., Xu, F., Datta, S. D., McQuillan, G. M., Berman, S. M., & Markowitz, L. E. (2009). Prevalence of sexually transmitted infections among female adolescents aged 14 to 19 in the United States. *Pediatrics*, *124*, 1505-1512.

Fulker, DW. Human genetics, part A: The unfolding genome (Progress in clinical and biological research 103A). New York: Alan R Liss; 1982. Extension of the classical twin method; p. 395-406.

Gallo, M. F., Steiner, M. J., Hobbs, M. M., Warner, L., Jamieson, D. J., & Macaluso, M. (2013). Biological markers of sexual activity: tools for improving measurement in HIV/sexually transmitted infection prevention research. *Sexually Transmitted Diseases*, *40*, 447-452.

Garrison, S.M., & Rogers, J. L. (2016). Casting doubt on the causal link between intelligence and age of first intercourse: A cross-generational sibling comparison design using the NLSY. *Intelligence, 59*, 139-156.

Geronimus, A. T., & Korenman, S. (1993). Maternal youth or family background? On the health disadvantages of infants with teenage mothers. *American Journal of Epidemiology*, *137*, 213-225.

Gesselman, A. N., Webster, G. D., & Garcia, J. R. (2017). Has virginity lost its virtue? Relationship stigma associated with being a sexually inexperienced adult. *The Journal of Sex Research*, *54*, 202-213.

Gipson, J. D., Koenig, M. A., & Hindin, M. J. (2008). The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. *Studies in Family Planning*, *39*, 18-38.

Goldberg, S. K., Haydon, A. A., Herring, A. H., & Halpern, C. T. (2014). Longitudinal consistency in self-reported age of first vaginal intercourse among young adults. *Journal of Sex Research*, *51*, 97-106.

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: a new perspective and a model for research. *Psychological Review*, *87*, 301-315.

Graham, C. A., Catania, J. A., Brand, R., Duong, T., & Canchola, J. A. (2003). Recalling sexual behavior: a methodological analysis of memory recall bias via interview using the diary as the gold standard. *Journal of Sex Research*, *40*, 325-332.

Guo, J., Chung, I. J., Hill, K. G., Hawkins, J. D., Catalano, R. F., & Abbott, R. D. (2002). Developmental relationships between adolescent substance use and risky sexual behavior in young adulthood. *Journal of Adolescent Health*, *31*, 354-362.

Guo, G., Tong, Y., Xie, C. W., & Lange, L. A. (2007). Dopamine transporter, gender, and number of sexual partners among young adults. *European Journal of Human Genetics*, *15*, 279-287.

Han, C., McGue, M., & Iacono, W. G. (1999). Lifetime tobacco, alcohol and other substance use in adolescence Minnesota twins: Univariate and multivariate behavioral genetic analyses. *Addiction, 94*, 981-993.

Halpern-Fisher, B.L., Millstein, S.G., & Ellen, J.M. (1996). Relationship of alcohol use and risky sexual behavior: A review and analysis of findings. *Journal of Adolescent Health, 19,* 331-336.

Harden, K. P. (2012). True love waits? A sibling-comparison study of age at first sexual intercourse and romantic relationships in young adulthood. *Psychological Science*, *23*, 1324-1336.

Harden, K. P. (2014). Genetic influences on adolescent sexual behavior: Why genes matter for environmentally oriented researchers. *Psychological Bulletin*, *140*, 434-465.

Harden, K. P., & Mendle, J. (2011a). Why don't smart teens have sex? A behavioral genetic approach. *Child Development*, *82*, 1327-1344.

Harden, K. P., & Mendle, J. (2011b). Adolescent sexual activity and the development of delinquent behavior: The role of relationship context. *Journal of youth and adolescence*, *40*, 825-838.

Harden, K. P., Mendle, J., Hill, J. E., Turkheimer, E., & Emery, R. E. (2008). Rethinking timing of first sex and delinquency. *Journal of Youth and Adolescence*, *37*, 373-385.

Haydon, A. A., Herring, A. H., & Halpern, C. T. (2012). Associations between patterns of emerging sexual behavior and young adult reproductive health. *Perspectives on Sexual and Reproductive Health*, *44*, 218-227.

Haydon, A. A., Herring, A. H., Prinstein, M. J., & Halpern, C. T. (2012). Beyond age at first sex: Patterns of emerging sexual behavior in adolescence and young adulthood. *Journal of Adolescent Health*, *50*, 456-463.

Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics*, *23*, 29-50.

Heldman, C., & Wade, L. (2010). Hook-up culture: Setting research agenda. *Sexuality Research and Social Policy, 7,* 323-333.

Henshaw, S. K. (1995). Factors hindering access to abortion services. *Family planning perspectives*, 54-87.

Heywood, W., Patrick, K., Smith, A. M., & Pitts, M. K. (2015). Associations between early first sexual intercourse and later sexual and reproductive outcomes: a systematic review of population-based data. *Archives of sexual behavior*, *44*, 531-569.

Hicks, B. M., Blonigen, D. M., Kramer, M. D., Krueger, R. F., Patrick, C. J., Iacono, W. G., & McGue, M. (2007). Gender differences and developmental change in externalizing disorders from late adolescence to early adulthood: A longitudinal twin study. *Journal of Abnormal Psychology*, *116*, 433-447.

Hicks, B.M., Krueger, R. F., Iacono, W. G., McGue, M., & Patrick, C. J. (2004). Family transmission and heritability of externalizing disorders: A twin-family study. *Archives of General Psychiatry, 61,* 922-928.

Hicks, B. M., Schalet, B. D., Malone, S. M., Iacono, W. G., & McGue, M. (2011). Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behavior Genetics*, *41*, 459-475.

Higgins, J. A., Mullinax, M., Trussell, J., Davidson Sr, J. K., & Moore, N. B. (2011). Sexual satisfaction and sexual health among university students in the United States. *American Journal of Public Health*, *101*, 1643-1654.

Hoffmann, T. J., Kvale, M. N., Hesselson, S. E., Zhan, Y., Aquino, C., Cao, Y., Cawley, S., Chung, E., Connell, S., Eshragh, J., Ewing, M. & Gollub, J., Henderson, M., Hubbell, E., Iribarren, C., Kaufman, J., Lao, R. Z., Lu, Y., Quesenberry, C., Ranatunga, D., Rowell, S., Sadler, M., Shapero, M. H., Shen, L., Shenoy, T. R., Smethurst, D., Van den Eeden, S. K., Walter, L., Wan, E., Wearley, R., Webster, T., Wen, C. C., Weng, L., Whitmer, R. A., Williams, A., Wong, S. C., Zau,C., Finn, A., Schaefer, C., Kwok, P., & Risch, N. (2011). Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*, *98*, 79-89.

[H]olway, G. V., Tillman, K. H., & Brewster, K. L. (2015). Binge drinking in young adulthood: the influence of age at first intercourse and rate of sex partner accumulation. *Archives of Sexual Behavior*, doi:10.1007/s10508-015-0597-y

Huibregtse, B. M., Bornovalova, M. A., Hicks, B. M., McGue, M., & Iacono, W. (2011). Testing the role of adolescent sexual initiation in later-life sexual risk behavior: A longitudinal twin design. *Psychological Science, 22,* 924-933.

Huibregtse, B. M., Corley, R. P., Wadsworth, S. J., Vandever, J. M., DeFries, J. C., & Stallings, M. C. (2016). A longitudinal adoption study of substance use behavior in adolescence. *Twin Research and Human Genetics: the official journal of the International Society for Twin Studies*, 1-11.

Iacono, W. G., Malone, S. M., & McGue, M. (2008). Behavioral disinhibition and the development of early-onset addiction: Common and specific influences. *Annual Review of Clinical Psychology, 4,* 325-348.

IBM Corp. Released 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, 21-37.

International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*, 52-58.

Jackson, C., Geddes, R., Haw, S., & Frank, J. (2012). Interventions to prevent substance use and risky sexual behavior in young people: A systematic review. *Addiction, 107,* 733-747.

Jessor, R., & Jessor, S. L. (1977). Problem behavior and psychosocial development: A longitudinal study of youth. San Diego, CA; Academic Press.

Kahn, J. A., Kaplowitz, R. A., Goodman, E., & Emans, S. J. (2002). The association between impulsiveness and sexual risk behaviors in adolescent and young adult women. *Journal of Adolescent Health*, *30*, 229-232.

Karlsson, R., Jonsson, M., Edlund, K., Evander, M., Gustavsson, Å., Boden, E., Rylander, E., & Wadell, G. (1995). Lifetime number of partners as the only independent risk factor for human papillomavirus infection: a population-based study. *Sexually transmitted diseases*, *22*, 119-127.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry, 60,* 929- 937.

Kendler, K. S., Schmitt, E., Aggen, S. H., & Prescott, C. A. (2008). Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Archives of General Psychiatry, 65,* 674-682.

Keyes, M. A., Malone, S. M., Elkins, I. J., Legrand, L. N., McGue, M., & Iacono, W. G. (2009). The enrichment study of the Minnesota twin family study: increasing the yield of twin families at high risk for externalizing psychopathology. *Twin Research and Human Genetics*, *12*, 489-501.

Kinsey, A.C., Pomeroy, W.B, & Martin, C.E. (1948). *Sexual Behavior of the Human Male.* Philadelphia, Pennsylvania: WB Saunders Co.

Kinsey, A.C., Pomeroy, W.B, Martin, C.E., & Gebhard, P.H. (1953). *Sexual Behavior of the Human Female: By the Staff of the Institute of Sex Research.* Philadelphia, Pennsylvania: WB Saunders Co.

Koblin, B. A., Husnik, M. J., Colfax, G., Huang, Y., Madison, M., Mayer, K., Barresi, P. J., Chesney, T. J., Buchbinder, M. A., & Buchbinder, S. (2006). Risk factors for HIV infection among men who have sex with men. *Aids*, *20*, 731-739.

Koopsman, J.R., Lorenz, J.P., & Boomsma, D. I. (1997). Association between alcohol use and smoking use in adolescent and young adult twins: A bivariate genetic analysis. *Alcoholism: Clinical and Experimental Research, 21,* 537-546.

Krueger, R. F., (1999). The structure of mental disorders. *Archives of General Psychiatry, 10,* 921-926.

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, *111*(3), 411.

Krueger, R. F. & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2,* 111–133.

Krueger, R. F. & Markon, K. E. (2011). A dimensional-spectrum model of psychopathology: Progress and opportunities. *Archives of General Psychiatry, 68,* 10-11.

Lang, A., (1985). The social psychology of drinking and human sexuality. *Journal of Drug Issues, 15*, 273– 289.

Lee, J. J., Vattikuti, S., & Chow, C. C. (2016). Uncovering the genetic architectures of quantitative traits. *Computational and Structural Biotechnology Journal*, *14*, 28-34.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, *88*, 294-305.

Leaity, S., Sherr, L., Wells, H., Evans, A., Miller, R., Johnson, M., & Elford, J. (2000). Repeat HIV testing: high-risk behaviour or risk reduction strategy?. *Aids*, *14*, 547-552.

Leigh, B.C., & Stall, R. (1993). Substance use and risky sexual behavior for exposure to HIV: Issues in methodology, interpretation, and prevention. *American Psychologist, 48,* 1035-1045.

Leigh B.C., Vanslyke, J.G., Hoppe, M.J., Rainey, D.T., Morrison, D.M., & Gillmore, M.R. (2008) Drinking and condom use: Results from an event-based daily diary. *AIDS Behavior, 12,* 104-112.

Lescano, C. M., Vazquez, E. A., Brown, L. K., Litvin, E. B., Pugatch, D., & Project SHIELD Study Group. (2006). Condom use with "casual" and "main" partners: What's in a name? *Journal of Adolescent Health*, *39*, DOI: http://dx.doi.org/10.1016/j.jadohealth.2006.01.003

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*, 1739-1740.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, and personality: A study of 850 sets of twins*. University of Texas Press, Austin.

Lopez-Leon, M., & Raley, J. (2012) Developmental Risks for Substance Use in Adolescence: Age as Risk Factor. *Clinical Handbook of Adolescent Addiction*, 132-138.

LoParo, D., & Waldman, I. (2014). Twins' rearing environment similarity and childhood externalizing disorders: A test of the equal environments assumption. *Behavior Genetics*, *44*, 606-613.

Lurie, M. N., & Rosenthal, S. (2010). Concurrent partnerships as a driver of the HIV epidemic in sub-Saharan Africa? The evidence is limited. *AIDS and Behavior*, *14*, 17-24.

Lyons, M. J., Koenen, K. C., Buchting, F., Meyer, J. M., Eaves, L., Toomey, R., Eisen, S. A., Goldberg, J., Faraone, S. V., Ban, R. J., Jerskey, B. A., & Tsuang, M. T.(2004). A twin study of sexual behavior in men. *Archives of Sexual Behavior*, *33*, 129-136.

Macaluso, M., Demand, M. J., Artz, L. M., & Hook III, E. W. (2000). Partner type and condom use. *Aids*, *14*, 537-546.

Maes, H. H., Sullivan, P. F., Bulik, C. M., Neale, M. C., Prescott, C. A., Eaves, L . J., & Kendler, K. S. (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine, 34,* 1251-1261.

Maes, H. H., Woodard, C. E., Murrelle, L., Meyer, J. M., Silberg, J. L., Hewitt, J. K., Rutter, M., et al. (1999). Tobacco, alcohol and drug use in eight- to sixteen-year-old twins: the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Studies on Alcohol*, *60*, 293–305.

Maher, B. (2008). The case of the missing heritability. *Nature*, *456*, 18-21.

Manchia, M., Cullis, J., Turecki, G., Rouleau, G. A., Uher, R., & Alda, M. (2013). The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PloS one*, *8*, e76295.

Manhart, L. E., Epstein, M., Bailey, J. A., Hill, K. G., Haggerty, K. P., & Catalano, R. F. (2016). HIV/Sexually Transmitted Infection Prevention Messaging: Targeting Root Causes of Sexual Risk Behavior. *Sexually transmitted diseases*, *43*, 71-77.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Wittemore, A.S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., & Vissher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*, 747-753.

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychological Bulletin*, *137*, 856-879.

Mather, K., & Jinks, J. L. (1977). Introduction to biometrical genetics. Ithaca, New York: Cornell University Press.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*, 635-649.

McGue, M., & Iacono, W. G. (2005). The association of early adolescent problem behavior with adult psychopathology. *American Journal of Psychiatry*, *162*(6), 1118-1124.

McGue, M., Iacono, W.G., & Kruger, R. F. (2006a). Minnesota Center for Twin and Family Research, *Twin Research and Human Genetics, 9,* 978-984.

McGue, M., Iacono, W.G, & Krueger, R. F. (2006b). The association of early adolescent problem behavior and adult psychopathology: a multivariate behavioral genetic prospective. *Behavioral Genetics, 36,* 591-602.

McGue, M., Malone, S., Keyes, M., & Iacono, W. G. (2014). Parent-offspring similarity for drinking: A longitudinal adoption study. *Behavior Genetics, 44,* 620-628.

McGue, M., Osler, M., Christensen, K. Causal Inference and observational research: The utility of twins (2010). *Perspectives on Psychological Science, 5,* 546-556.

McGue, M., Sharma, A., & Benson, P. (1996). Parent and sibling influences of adolescent alcohol use and misuse: Evidence from a U.S. adoption cohort. *Journal of Studies on Alcohol, 57,* 8-18.Neale, M., & Cardon, L. (1992). *Methodology for genetic studies of twins and families*. Springer Science & Business Media.

McGue, M., Zhang, Y., Miller, M. B., Basu, S., Vrieze, S., Hicks, B., Malone, S., Oetting, W.S., & Iacono, W. G. (2013). A genome-wide association study of behavioral disinhibition. *Behavior Genetics*, *43*, 363-373.

McCarthy, M. I., & Hirschhorn, J. N. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Human Molecular Genetics*, *17*, R156-R165.

McKie, L., Al‐Bashir, M., Anagnostopoulou, T., Csepe, P., Fonseca, H., El‐Asfahani, A., Funiak, S., Javeta, R. & Samsuridjal, S. (1993). Defining and assessing risky behaviours. *Journal of Advanced Nursing*, *18*(12), 1911-1916.

Medland, S. E., & Keller, M. C. (2009). Modeling extended twin family data II: Power associated with different family structures. *Twin Research and Human Genetics*, *12*, 19-25.

Meier, A. M. (2007). Adolescent first sex and subsequent mental health. *American Journal of Sociology*, *112*, 1811-1847.

Meier, K. J., Haider-Markel, D. P., Stanislawski, A. J., & McFarlane, D. R. (1996). The impact of state-level restrictions on abortion. *Demography*, *33*, 307-312.

Mendle, J., Harden, K. P., Turkheimer, E., Van Hulle, C. A., D'onofrio, B. M., Brooks‐Gunn, J., Rodgers, J. L., Emery, R. E., & Lahey, B. B. (2009). Associations between father absence and age of first sexual intercourse. *Child Development*, *80*, 1463-1480.

Mercer, C. H. (2010) Measuring sxual behavior and risk. Retrieved from http://surveynet.ac.uk/sqb/topics/healthbehaviour/sqb_sex_mercer.pdf

Miller, M. B., Basu, S., Cunningham, J., Eskin, E., Malone, S. M., Oetting, W. S., Schork, N., Sul, J. H., Iacono, W. G. & McGue, M. (2012). The Minnesota Center for Twin and Family Research genome-wide association study. *Twin Research and Human Genetics*, *15*, 767-774.

Mirzaei, M., Ahmadi, K., Saadat, S. H., & Ramezani, M. A. (2016). Instruments of high risk sexual behavior assessment: A systematic review. *Materia socio-medica*, *28*, 46- 50.

Morris, M. (1993). Telling tails explain the discrepancy in sexual partner reports. *Nature*, *365*, 437-440.

Morris, M., & Kretzschmar, M. (1997). Concurrent partnerships and the spread of HIV. *AIDS, 11,* 641-648.

Morrison, D.M., Gillmore, R., Hoppe, M.J., Gaylord, J., Leigh, B.C., & Rainey, D. (2008). Adolescent drinking and sex: Findings from a daily diary study, *Perspectives on Sexual and Reproductive Health, 35,* 162-168.

Mustanski, B. S. (2001). Getting wired: Exploiting the Internet for the collection of valid sexuality data. *Journal of Sex Research*, *38*, 292-301.

Mustanski, B., Viken, R. J., Kaprio, J., Winter, T., & Rose, R. J. (2007). Sexual behavior in young adulthood: A population-based twin study. *Health Psychology*, *26*, 610- 617.

Muthén, L. K., & Muthén, B. O. (1998-2011). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Neale, M.C. (1997). Mx: Statistical modeling. Medical College of Virginia, Richmond.

Neale, M.C., & Cardon, L. (1992). *Methodology for genetic studies of twins and families* (Vol. 67). Springer Science & Business Media.

Neale, M.C., Duffy, D.L. & Martin, N.G. (1994). Direction of causation: reply to commentaries, *Genetic Epidemiology, 11,* 463–472.

Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*, *24*, 239-258.

Nelson, S. J., Manhart, L. E., Gorbach, P. M., Martin, D. H., Stoner, B. P., Aral, S. O., & Holmes, K. K. (2007). Measuring sex partner concurrency: it's what's missing that counts. *Sexually Transmitted Diseases*, *34*, 801-807.

Nichols, R. C., & Bilbro Jr, W. C. (1966). The diagnosis of twin zygosity. *Human Heredity*, *16*, 265-275.

Nicholson, N., Soane, E., Fenton-O'Creevy, M., & William, P. (2005). Personality and Domain-Specific Risk Taking. *Journal of Risk Research*, 8, 157-176

Noar, S. M., Cole, C., & Carlyle, K. (2006). Condom use measurement in 56 studies of sexual risk behavior: review and recommendations. *Archives of Sexual Behavior*, *35*(3), 327-345.

O'Hara, R. E., & Cooper, M. L. (2015). Bidirectional associations between alcohol use and sexual risk-taking behavior from adolescence into young adulthood. *Archives of Sexual Behavior*, *44*, 857-871.

Oshri, A., Tubman, J. G., Morgan‐Lopez, A. A., Saavedra, L. M., & Csizmadia, A. (2013). Sexual Sensation Seeking, Co‐occurring Sex and Alcohol Use, and Sexual Risk Behavior among Adolescents in Treatment for Substance Use Problems. *The American Journal on Addictions*, *22*(3), 197-205.

Owusu-Edusei, K. Jr., Chesson, H.W., Gift, T.L., Tao, G., Mahajan, R. Ocfemia, M.C.B., & Kent, C.K. (2013). The estimated direct medical cost of selected transmitted infections in the United States, 2008. *Sexually Transmitted Diseases, 40,* 197-201.

Pagan, J., Rose, R., Viken, R., & Pulkkinen, L. (2006). Genetic and environmental influences on stages of alcohol use across adolescence and into young adulthood. *Behavior Genetics*, *36*, 483–497.

Patel, P., Borkowf, C. B., Brooks, J. T., Lasry, A., Lansky, A., & Mermin, J. (2014). Estimating per-act HIV transmission risk: a systematic review. *Aids*, *28*, 1509-1519.

Paul, C., Fitzjohn, J., Herbison, P., & Dickson, N. (2000). The determinants of sexual intercourse before age 16. *Journal of Adolescent Health*, *27*, 136-147.

Petrill, S. A., Plomin, R., DeFries, J. C., & Hewitt, J. K. (Eds.). (2003). *Nature, nurture, and the transition to early adolescence*. Oxford University Press, USA

Plomin, R., & DeFries, J. C. (1983). The Colorado adoption project. *Child Development, 54,* 276-289.

Plomin, R., & DeFries, J. C. (1985). A parent-offspring adoption study of cognitive abilities in early childhood. *Intelligence*, *9*, 341-356.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559-575.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramrakha, S., Bell, M. L., Paul, C., Dickson, N., Moffitt, T. E., & Caspi, A. (2007). Childhood behavior problems linked to sexual risk taking in young adulthood: A birth cohort study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1272-1279.

Rhea, S. A., Bricker, J. B., Corley, R. P., DeFries, J. C., & Wadsworth, S. J. (2013). Design, Utility, and History of the Colorado Adoption Project: Examples Involving Adjustment Interactions. *Adoption Quarterly*, *16*, 17-39.

Rhea, S. A., Bricker, J. B., Wadsworth, S. J., & Corley, R. P. (2013b). The Colorado Adoption Project. *Twin Research and Human Genetics*, *16*, 358-365.

Rhea, S. A., Gross, A., Haberstick, B. C., & Corley, R. P. (2006). Colorado Twin Registry, *Twin Research and Human Genetics, 9,* 941-949.

Rhea, S. A., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2013a). Colorado twin registry: an update. *Twin Research and Human Genetics*, *16*, 351-357.

Rhee, S. H., Hewitt, J. K., Corley, R. P., Willcutt, E.G., & Pennington, B.F. (2005). Testing hypotheses regarding the causes of comorbidity: examining the underlying deficits of comorbid disorders. *Journal of Abnormal Psychology, 114,* 346-362.

Rhee, S. H., Hewitt, J. K., Young, S. E., Corley, R. P., Crowley, T. J., & Stallings, M. C. (2003). Genetic and environmental influences on substance initiation, use, and problem use in adolescents. *Archives of General Psychiatry, 60*, 1256-1264.

Rhem, J., Shield, K.D., Joharchi, N., & Shuper, P.A. (2012). Alcohol consumption and the intention to engage in unprotected sex: Systematic review and meta-analysis of experimental studies. *Addiction, 107,* 51-5

Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A., Esko, T., Medland, S. E., Vinkhuyzen, A. A. E., Yang, J., Boardman, J. D., Chabris, C. F., Daws,

C. T., Domnigue, B. W., Hinds, D. A., Johannesson, M., Kiefer, A. K., Laibson, D., Magnusson, P. K. E., Mountain, J. L., Oskarsson, S., Rostapshova, O., Teumer, A., Tung, J. Y., Vissher, P. M., Benjamin, D. J., Cesarini, D., Koellinger, P. D., & the Social Science Genetics Association Consortium (2014). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological Science*, *25*, 1975-1986.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467-1471.

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*, 421-427.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., Farmer, A., Jablenski, A., Pickens, R., Regier, D. A., Sartorius, N., Towle, L. H., (1988). The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic system and in different cultures. *Archives of General Psychiatry, 45,* 1069-1077.

Rose, R.J., Dick, D.M., Viken, R.J., Pulkkine, L., & Kaprio, J. (2001). Drinking or abstaining at age 14? A genetic epidemiological study. *Alcoholism: Clinical and Experimental Research, 11,* 1594-1604.

Rowe, D. C., Rodgers, J. L., Meseck-Bushey, S., & St John, C. (1989). Sexual behavior and nonsexual deviance: A sibling study of their relationship. *Developmental Psychology*, *25*, 61-69.

Rutter, M. (2007). Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science*, *2*, 377-395.

Ryan, S., Franzetta, K., Manlove, J. S., & Schelar, E. (2008). Older sexual partners during adolescence: links to reproductive health outcomes in young adulthood. *Perspectives on Sexual and Reproductive Health*, *40*, 17-26.

Sandfort, T. G., Orr, M., Hirsch, J. S., & Santelli, J. (2008). Long-term health correlates of timing of sexual debut: Results from a national US study. *American Journal of Public Health*, *98*, 155-161.

Santelli, J. S., & Beilenson, P. (1992). Risk factors for adolescent sexual behavior, fertility, and sexually transmitted diseases. *Journal of School Health*, *62*, 271-279.

Santelli, J. S., Brener, N. D., Lowry, R., Bhatt, A., & Zabin, L. S. (1998). Multiple sexual partners among US adolescents and young adults. *Family Planning Perspectives, 30*, 271-275.

Samek, D. R., Iacono, W. G., Keyes, M. A., Epstein, M., Bornovalova, M. A., & McGue, M. (2014). The developmental progression of age 14 behavioral disinhibition, early age of

sexual initiation, and subsequent sexual risk‐taking behavior. *Journal of Child Psychology and Psychiatry*, *55*, 784-792.

Sarason, I. G., Johnson, J. H., & Siegel, J. M. (1978). Assessing the impact of life change: Development of the Life Experience Survey. *Journal of Consulting and Clinical Psychology, 46,* 932-946.

Satterwhite, C.L., Torrone, E., Meites, E., Dunne, E.F., Mahajan, R., Ocfemia, M.C.B., Su, J., Xu, F., & Weinstock H. (2013). Sexually transmitted infections among US women and men: Prevalence and incidence estimates, 2008. *Sexually Transmitted Diseases, 40,* 187-193.

Schroder, K. E., Carey, M. P., & Vanable, P. A. (2003a). Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytical options. *Annals of Behavioral Medicine*, *26*, 76-103.

Schroder, K. E., Carey, M. P., & Vanable, P. A. (2003b). Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. *Annals of Behavioral Medicine*, *26*, 104-123.

Sher, K. J., & Trull, T. J. (1994). Personality and disinhibitory psychopathology: Alcoholism and antisocial personality disorder. *Journal of Abnormal Psychology, 103*, 92–102.

Shuper, P.A., Neuman, M., Kanteres, F., Baliunas, D., Joharchi, N., & Rhem, J. (2010). Causal considerations on alcohol and HIV/AIDS – A systematic review. *Alcohol & Alcoholism, 45,* 159-166.

Spriggs, A. L., & Halpern, C. T. (2008). Sexual debut timing and depressive symptoms in emerging adulthood. *Journal of Youth and Adolescence*, *37*, 1085-1096.

Stallings, M. C., Corley, R. P., Dennehey, B., Hewitt, J. K., Krauter, K. S., Lessem, J. M., Mikulick-Gilbertson, S. K., Rhee, S. H., Smolen, A., Young, S. E., & Crowley, T. J. (2005). A genome-wide search for quantitative trait loci that influence antisocial drug dependence in adolescence. *Archives of General Psychiatry*, *62*, 1042-1051.

Stallings, M.C., Corley, R.P., Hewitt, J.K., Krauter, K.S., Lessem, J.M., Mikulich, S.K., Rhee, S.H., Smolen, A., Young, S.E., & Crowley, T.J. (2003). A genome-wide search for quantitative trait loci influencing substance dependence vulnerability in adolescence. h*Drug and Alcohol Dependence*, *70*, 295-307.

Stallings, M. C., Gizer, I. R., & Young-Wolff, K. C. (2016). Genetic epidemiology and molecular genetics. *The Oxford Handbook of Substance Use and Substance Use Disorders*. Retrieved from DOI: 10.1093/oxfordbb/9780199381678.013.002

Steele, C.M., & Josephs, R.A., (1990). Alcohol myopia: Its prized and dangerous effects. *American Psychologist, 45,* 921-933.

Stringer, S., Minică, C. C., Verweij, K. J. H., Mbarek, H., Bernard, M., Derringer, J., et al. (2016). Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Translational Psychiatry*, *6*, e769.

So, H. C., & Sham, P. C. (2016). Exploring the predictive power of polygenic scores derived from genome-wide association studies: a study of 10 complex traits. *Bioinformatics*, *33*, 886-892.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, *14*, 483-495.

Sturdevant, M. S., Belzer, M., Weissman, G., Friedman, L. B., Sarr, M., Muenz, L. R., & Adolescent Medicine HIV/AIDS Research Network. (2001). The relationship of unsafe sexual behavior and the characteristics of sexual partners of HIV infected and HIV uninfected adolescent females. *Journal of Adolescent Health*, *29*, 64-71.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, *12*, e1001779.

Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*, *42*, 441-447.

Turchik, J. A., & Garske, J. P. (2009). Measurement of sexual risk taking among college students. *Archives of Sexual Behavior*, *38*, 936-948.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science*, *280*, 867-873.

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2015). Changes in American adults' sexual behavior and attitudes, 1972–2012. *Archives of Sexual Behavior*, *44*, 2273-2285.

Valleroy, L. A., MacKellar, D. A., Karon, J. M., Rosen, D. H., McFarland, W., Shehan, D. A., Stoyanoff, S. R., LaLota, M., Celentano, D. D. Koblin, B. A., Theide, H., Katz, M. H., Torian, L. V., Janssen, R. S., & the Young Men's Survey Study Group (2000). HIV prevalence and associated risks in young men who have sex with men. *JAMA*, *284*, 198-204.

Van Duynhoven, Y. T., Nagelkerke, N. J., & Van De Laar, M. J. (1999). Reliability of Self-Reported Sexual Histories: Test-Retest and Interpartner Comparison in a Sexually Transmitted Diseases Clinic. *Sexually Transmitted Diseases*, *26*, 33-42.

Vasilenko, S. A., Kugler, K. C., Butera, N. M., & Lanza, S. T. (2015a). Patterns of adolescent sexual behavior predicting young adult sexually transmitted infections: A latent class analysis approach. *Archives of Sexual Behavior*, *44*, 705-715.

Vasilenko, S. A., Kugler, K. C., & Lanza, S. T. (2016). Latent classes of adolescent sexual and romantic relationship experiences: Implications for adult sexual health and relationship outcomes. *The Journal of Sex Research*, *53*, 742-753.

Vitaro F, Brendgen M, Arseneault L. (2009). The discordant MZ-twin method: One step closer to the holy grail of causality. *International Journal of Behavioral Development, 33,* 376–382.

Verweij, K. J. H., Zietsch, B. P., Bailey, J. M., & Martin, N. G. (2009). Shared aetiology of risky sexual behaviour and adolescent misconduct: Genetic and environmental influences. *Genes, Brain and Behavior*, *8*, 107-113.

Viken, R.J., Kaprio, J., Koskenvuo, M., & Rose, R. J. (1999). Longitudinal analysis of the determinates of drinking and of drinking to intoxication in adolescent twins. *Behavior Genetics, 29,* 455-461.

Vrieze, S. I., McGue, M, Miller, M B., Hicks, B. M., & Iacono, W. G. (2013). Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/ dependence, and their co-occurrence: Twin biometry GCTA, and genome-wide scoring. *Behavior Genetics, 43,* 97-107.

Walsh, J.L., Fielder, R.L., Carey, K.B., & Carey, M.P. (2013). Do alcohol and marijuana use decrease the probability of condom use for college women? *Journal of Sex Research, 51,* 145-158.

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). FUMA: Functional mapping and annotation of genetic associations. *bioRxiv*, 110023.

Wosick-Correa, K. (2010). Agreements, rules and agentic fidelity in polyamorous relationships. *Psychology & Sexuality*, *1*, 44-61.

Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R. Ripke, S., MacIntyre, D. J., McGhee, K.A., Maclean A.W., Smit, J.H. Hottenga, J.J., Willemsen, G., Middeldorp, C.M., de Geus, E.J.C., Lewis, C.M., McGuffin, P., Hickie, I.B., van de Oord, E.J.C.G., Liu, J.Z., Macgregor, S., McEvoy, B.P., Byrne, E.M., Medland, S.E., Statham, D.J., Henders, A.K., Health, A.C., Montgomery, G.W., Martin, N.G., Boomsma, D.I., Madden, P.A.F., & Sullivan, P.F. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Molecular Psychiatry*, *17*, 36-48.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*, 507- 515.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Health, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Vissher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*, 565-569.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*, 76-82.

Young, M., Palacios, R., & Penhollow, T. M. (2012). Measurement and design issues in the study of adolescent sexual behavior and the evaluation of adolescent sexual health behavior interventions. *American Journal of Sexuality Education*, 429-444.

Young, S.E., Rhee, S.H., Stallings, M.C., Corley, R.P., & Hewitt, J. K. (2006). Genetic and environmental vulnerabilities underlying adolescent substance use and problem use: General or Specific? *Behavioral Genetics, 603-615.*

Young, S. E., Stallings, M. C., Corley, R. P., Krauter, K. S., & Hewitt, J. K. (2000). Genetic and environmental influences on behavioral disinhibition. *American journal of medical genetics*, *96*, 684-695.

Zenilman, J. M., Weisman, C. S., Rompalo, A. M., Ellish, N., Upchurch, D. M., Hook III, E. W., & Celentano, D. (1995). Condom use to prevent incident STDs: the validity of self-reported condom use. *Sexually Transmitted Diseases*, *22*, 15-21.

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., Hemani, G., Tansey, K., Laurin, C., Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, St Pourcain, B., Warrigton, N. M., Finucane, H. K., Price, A. L., Bulik-Sullivan, B. K., Anttila, V., Paternoster, L., Gaunt, T. R., Evans, D. M., & Neale, B. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*, 272-279.

Zietsch, B. P., Morley, K. I., Shekar, S. N., Verweij, K. J., Keller, M. C., Macgregor, S., Wright, M. J., Bailey, J. M., & Martin, N. G. (2008). Genetic factors predisposing to homosexuality may increase mating success in heterosexuals. *Evolution and Human Behavior*, *29*, 424-433.

Zietsch, B. P., Verweij, K. J. H., Bailey, J. M., Wright, M. J., & Martin, N. G. (2010). Genetic and environmental influences on risky sexual behaviour and its relationship with personality. *Behavior Genetics*, *40*, 12-21.

Zimmer-Gembeck, M. J., & Helfand, M. (2008). Ten years of longitudinal research on US adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background. *Developmental Review*, *28*, 153-224.

Zolna, M., & Lindberg, L. D. (2012). *Unintended pregnancy: Incidence and outcomes among young adult unmarried women in the United States, 2001 and 2008*. New York: Alan Guttmacher Institute.

APPENDIX A

Modified Risk Behavior Questionnaire (Short-MRBQ)

Module A: Sexual Behavior

Note to reviewers: Module A will be self-administered on a secure website or laptop computer, depending on whether subjects are tested via telephone or in-person; items are presented one-at-a-time so that automatic skip patterns will prevent some items being viewed; numerical responses will be recorded via a pull-down menu (options include <10 years, 10 years, 11 years, etc., up to highest age of participants); a "prefer not to answer" option will be available for all items; the computer will navigate all skip patterns automatically.

****

Now we would like to ask you some questions about your sexual behavior. These questions will take most people about 5 minutes to answer. Because different behaviors may have different health risks, we include questions about specific sexual activities. Your answers to these questions will help us to understand how these behaviors are related to health. We realize that the following questions are very personal. Your answers will remain strictly confidential and will be associated only with a numerical ID code, not your name. However, if you would prefer not to answer an item, please indicate that preference using the option provided, and move on to the next question.

1. Are you? () Male () Female

2. Are you? () Not Married () Married, living together () Separated/divorced () Widow/ Widower

3. In your lifetime, with how many people (different sex partners) have you had oral, vaginal, or anal sex?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

(If None, skip to end of Module A).

4a. In your lifetime, how many of your sex partners were female?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

4b. In your lifetime, how many of your sex partners were male?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

5. How old were you the *first* time you had oral sex? __ __
(**Never had oral sex is an option—skip to 6**).

5a. How old was your *partner* the first time you had oral sex? __ __

5b. How old were you the first time you *willingly agreed* to oral sex? __ __ (“**Same as above” is an option)**

5c. How old was your partner then? __ __

6. How old were you the *first* time you had vaginal sex? __ __
(Never had vaginal sex is an option -skip to 7).

6a. How old was your *partner* the first time you had vaginal sex? __ __

6b. how old were you the first time you *willingly agreed* to vaginal sex? __ __

6c.How old was your partner then? __ __

7. How old were you the *first* time you had anal sex? __ __
(Never had anal sex is an option—skip to 8).

7a. How old was your *partner* the first time you had anal sex? __ __

7b. How old were you the first time you *willingly agreed* to anal sex? __ __

7c. How old was your partner then? __ __

****

Now we would like to ask you some questions your sexual behavior in the *last 12 months (past year)*.

8. Have you engaged in sexual activity in the last 12 months? () Yes () No

(If No, skip to item 31 and score item 9 as “none”)

9. In the past 12 months, with how many people (different sex partners) have you had oral, vaginal, or anal sex?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

10. In the past 12 months, how many of these sex partners would you consider a regular partner (with whom you have, or had a relationship)?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

11. In the past 12 months, how many of these sex partners would you consider a casual partner (with whom you did not have a relationship)?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+

9. In the last 12 months, alcohol or drugs has influenced my decision to do something sexual with a partner.

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

10. In the last 12 months, I have used alcohol or drugs to help me feel more comfortable with a sexual partner.

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

11. In the last 12 months, I have done more sexually with a partner than I planned because I was drinking or
using drugs.

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

12. In the last 12 months, I have had unprotected sex (not used a condom) because I was drinking or using
drugs.

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

13. In the last 12 months, I have regretted having sex with someone because I was using drugs or alcohol.

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

****

Now we would like to ask you some questions about your condom use.

14a. Recall the last time you had unprotected sex and DID NOT use a condom for protection. Were you using alcohol or drugs at the time?

() No () Yes () I always use condoms so I can't answer this question yes or no

14b. Recall the last time you had sex and DID use a condom for protection.
Were you using alcohol or drugs at the time?

() No () Yes () I never use condoms so I can't answer this question yes or no

16. In the last 12 months, how often did you use a condom when you had oral sex?

() Never () Less than half the time () About half the time () More than half the time () Always

() I don't, or didn't have oral sex in the past 12 months

(If scored always or don't, or didn't have oral sex….), skip to 16)

16. In the last 12 months, how often did you use a condom when you had vaginal sex?

() Never () Less than half the time () About half the time () More than half the time () Always

() I don't, or didn't have vaginal sex in the past 12 months

(If scored always or don't, or didn't have vaginal sex….), skip to 17)

16b. During those times when you did not use condoms, were you in a regular (exclusive) relationship?

()Yes ()No

17. In the last 12 months, how often did you use a condom when you had anal sex?

() Never () Less than half the time () About half the time () More than half the time () Always

() I don't, or didn't have anal sex in the past 12 months

(If scored always or don't or didn't have anal sex….) skip to 18)

17b. During those times when you did not use condoms, were you in a regular (exclusive) relationship?

()Yes ()No

****

Now we would like to ask you some questions about your sexual partners in the last 12 months.

18. In the last 12 months, how many times have you had sex with someone you didn't know well or just met?

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

19. In the last 12 months, how many times have you had sex with a new partner before discussing their sexual history and risk for sexually transmitted diseases (STDs)?

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

21. In the last 12 months, how many sex partners (that you know of) have you had sex with who had been sexually active before you but had not been tested for sexually transmitted diseases (STDs) and HIV?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

22. In the last 12 months, how many sex partners have you had sex with that you didn't trust?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

23. In the last 12 months, how many of your sex partners were having sex with someone other than you?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

24. In the last 12 months, how many of your sex partners were injection drug users?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

25. In the last 12 months, how many of your sex partners were HIV positive or had AIDS?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

26. In the last 12 months, how many of your sex partners had hepatitis?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

27. In the last 12 months, how many of your sex partners had an (STD) like chlamydia, HPV, gonorrhea or syphilis?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

28. In the past 12 months, how many of your sex partners had herpes?

() None () 1 () 2 () 3–5 () 6–9 () 10–19 () 20+ () I don't know

****

Finally, we would like to ask you some questions about your health.

31. Have you ever been told by a doctor, nurse, or other health care professional that you had:

a. Hepatitis …….	() No	() Yes
b. Gonorrhea……..	() No	() Yes
c. Syphilis……..	() No	() Yes
d. HPV……	() No	() Yes
e. Chlamydia (or NGU)….	() No	() Yes
f. HIV or AIDS	() No	() Yes
g. Herpes.....….	() No	() Yes

32. **(If male):** Have you ever gotten a partner pregnant unintentionally? () No () Yes

32a. How many times __ __? **(Pull down menu, 1 to 10+)?**

33. **(If female):** Have you ever had an unintended pregnancy? () No () Yes

33b. How many times __ __? **(Pull down menu, 1 to 10+)?**

34. Have you ever, in your lifetime, given sex to get money?

() No **(If No, skip 34a)** () Yes

34a. In the last 12 months, how many times did you give sex to get money?

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times
() 20+ times

35. Have you ever, in your lifetime, given sex to get drugs?

() No **(If No, skip 35a)** () Yes

35a. In the last 12 months, how many times did you give sex to get drugs?

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

34. Given your present behavior and partners, what do you think your chances of getting HIV is?

() No chance () Some chance () 50-50 chance () High chance () Sure chance

35. Do you consider yourself to be:

() 100% heterosexual (straight)
() mostly heterosexual (straight), but somewhat attracted to people of your own sex
() bisexual—that is, attracted to men and women equally
() mostly homosexual (gay or lesbian), but somewhat attracted to people of the opposite sex
() 100% homosexual (gay or lesbian)
() not sexually attracted to either males or females

36. How satisfied are you with your sex life?

() Terribly dissatisfied
() Quite dissatisfied
() Somewhat dissatisfied
() Neutral
() Somewhat satisfied
() Quite satisfied
() Delighted

37. Is there anything else we should know that may help us understand your responses? For example, were you unable to have sex during the past year due to an illness or injury? Or, due to military service, or your job, were you away from your sex partner for much of the past year?

(text box for open response)

END OF MODULE A

****

APPENDIX B

College Behavior Survey

Module A: Sexual Behavior (see previous Appendix)

Module B: Drug Use and Risk Behavior

****

This survey concerns sexual behavior and drug use. We understand that some of the questions may be sensitive in nature and we would like to assure you that your responses will be completely anonymous. There will be no attempt to identify single participants, rather researchers will be looking at trends across the whole sample.

Now we would like to know about your recreational marijuana use.

38. Have you used marijuana in the last 12 months?

() No () Yes

39. In the last 12 months, you have used marijuana:

() Less than once a month () Monthly () Weekly () Daily or near daily

40. On days which you use marijuana, what proportion of the hours that you are awake are spend doing activities associated with consuming marijuana (for example, smoking, vaporizing, eating marijuana, time spent high)?

__ __ (0-100%, in 10% increments)

41. In which ways do you use marijuana (Select all that apply):

() Smoking (ex: joins, pipes, bongs)
() Vaporizers
() Edibles, Tinctures/Tonics, Marijuana Drinks
() Hash, Wax, or Dabs

42. What is your preferred method to use marijuana?
() Smoking (ex: joins, pipes, bongs)
() Vaporizers
() Edibles, Tinctures/Tonics, Marijuana Drinks
() Hash, Wax, or Dabs

43. During the last year, in a typical month, how often did you use/do the following:

a. Smoking
() Daily () 4-6 days a week () 2-3 days a week () once a week
() 2-3 times a month () Once a month or less () Never

b. Vaporizers
() Daily () 4-6 days a week () 2-3 days a week () once a week
() 2-3 times a month () Once a month or less () Never

c. Edibles, Tinctures/ Tonics, Drinks
() Daily () 4-6 days a week () 2-3 days a week () once a week
() 2-3 times a month () Once a month or less () Never

d. Hash, Wax, Dabs
() Daily () 4-6 days a week () 2-3 days a week () once a week
() 2-3 times a month () Once a month or less () Never

44. Since recreational marijuana became legal in Colorado, do you think your typical use has:

() Increased a lot () Increased a little () Stayed the same
() Decreased a little () Decreased a lot

45. Is there anything else you could tell us about your typical marijuana use pattern that may clarify your responses?

(text box for open response)

46. Are there any comments about our marijuana questionnaire that may help us improve it in the future (for example, confusing working, mistakes, or suggestions)?

(text box for open response)

****

We asked you before about how your drug and alcohol use influenced your sexual behavior. Now we would like to know if your marijuana use has influenced your sexual behavior.

47. In the past 12 months did:

a. Marijuana influence a decision to do something sexual with a partner
() No () Yes

b. You use marijuana to feel more comfortable with a sexual partner

() No () Yes

c. You do more sexually because you were using marijuana

() No () Yes

d. You have unprotected sex because you were using marijuana

() No () Yes

e. You regret having sex because you were using marijuana

() No () Yes

****

48. In the last 12 months, how often were you using marijuana when you drove a vehicle:

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

49. In the last 12 months, how often were you using marijuana when you went to work:

() Never () 1 time () 2 times () 3–5 times () 6–9 times () 10–19 times () 20+ times

****

50. a. In the past year, have you engaged in activities with water skiing, surfing, or scuba diving?

() Never () Rarely () Sometimes () Often () Very Often

b. When you engage in these activities, do you make sure to use the recommended safety procedures and protective equipment?

() Never () Rarely () Sometimes () Often () Very Often

c. Have you engaged in these activities while using alcohol or drugs?

() Never () Rarely () Sometimes () Often () Very Often

d. When you do these activities do you like to push yourself to the limits or risk your personal safety?

() Never () Rarely () Sometimes () Often () Very Often

51. a. In the past year, have you ridden a motorcycle (street bike), dirt bike, or ATV?

() Never () Rarely () Sometimes () Often () Very Often

b. When you ride, do you wear a helmet?

() Never () Rarely () Sometimes () Often () Very Often

c. Have you engaged in these activities while using alcohol or drugs?

() Never () Rarely () Sometimes () Often () Very Often

d. When you do these activities do you like to push yourself to the limits or risk your personal safety?

() Never () Rarely () Sometimes () Often () Very Often

52. a. In the past year, have you engaged in activities like snowmobiling, jet skiing, or boat racing?

() Never () Rarely () Sometimes () Often () Very Often

b. When you engage in these activities, do you make sure to use the recommended safety procedures and protective equipment?

() Never () Rarely () Sometimes () Often () Very Often

c. Have you engaged in these activities while using alcohol or drugs?

() Never () Rarely () Sometimes () Often () Very Often

d. When you do these activities do you like to push yourself to the limits or risk your personal safety?

() Never () Rarely () Sometimes () Often () Very Often

53. a. In the past year, have you engaged in activities like rock climbing, sky diving, or hang gliding?

() Never () Rarely () Sometimes () Often () Very Often

b. When you engage in these activities, do you make sure to use the recommended safety procedures and protective equipment?

() Never () Rarely () Sometimes () Often () Very Often

c. Have you engaged in these activities while using alcohol or drugs?

() Never () Rarely () Sometimes () Often () Very Often

d. When you do these activities do you like to push yourself to the limits or risk your personal safety?

() Never () Rarely () Sometimes () Often () Very Often

****

We are interested in why you selected a previous item. If nothing describes you well, please give us some feedback to make our survey more accurate and inclusive.

54. You answered that alcohol and drugs has sometimes influenced your decision to do something sexual with a partner. Which of the following best describes what you meant:

() You made a decision to sleep or hook up with **someone you would NOT have chosen in sober**

() If you think back, **you may or may not have had sex or hooked up with this person if you were sober,** but the decision was made easier (choose to) because you were impaired

() You happened to be drug or high a situation leading up to se or hooking up, but your **decision would probably be the same if you were sober**

() Other, please explain

55. You answered that you sometimes use drugs and alcohol to feel more comfortable during sex. Which of the following best describes what you meant:

() to reduce physical discomfort associated with sex

() to feel less shy **around my partner**

() to feel less shy **about specific sexual requests or desires**

() to feel more turned on to get in the mood

() to feel comfortable with my decision to have sex

() Other, please explain

56. You answered that you have sometimes done more sexually with a partner than planned because you were drinking or using drugs. Which of the following best describes what you meant:

You had not previously planned on having sexual contact with the person or within a specific situation, but did:

() You were happy with the outcome

() You regretted the outcome

() Other, please explain

You were hoping to have sexual contact with the person or in a specific situation, but went further than planned:

() You were happy with the outcome

() You regretted the outcome

() Other, please explain

57. You answered that you have had unprotected sex (not used a condom) because you were drinking or using drugs. There may be a lot of reasons why you had unprotected sex, please describe which **best** suits your typical behavior:

I had unprotected sex at times, where I would have had protected sex when sober:

() I typically use condoms sober, but not when impaired

() Other, please explain

I had unprotected sex at times, where I would have used a condom sober:

() It was too much trouble in the moment to get a condom

() Using a condom did not cross my mind, or I forgot

() Other, please explain

I had unprotected sex at times, but it would have been unprotected when sober too:

() I never wear or use condoms for any reason because I don't like them

() I never wear or use condoms because pregnancy is not a concern (i.e. other forms of birth control, do not have heterosexual vaginal sex, other)

() I never wear or use condoms because I am not concerned about disease (i.e. trustworthy regular partner)

() I am neither concerned about pregnancy or disease (for reasons listed above or other)

() Other, please explain

58. You answered that you have sometimes regretted as sexual encounter because you were drinking or using drugs. Which of the following best describes what you meant:

() I would not have had sex sober, I wish I hadn't with that person or at that time

() There were some social consequences related to those encounters (i.e., I am embarrassed, hurt my relationship with someone or a partner

() There were physical consequences related to those encounters (i.e., unintended pregnancy, new STD)

() Other, please explain

59. Is there anything else we should know that may help us understand your responses (for example, "due do an chronic illness, I was unable to have sex for the past year")?

(text box for open response)

60. Are there any final comments about our survey that may help us improve it in the future (for example, confusing working, mistakes, or suggestions)?

(text box for open response)