



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations


Graduate School

2017

The Influence of Genetic and Environmental Factors on Quit Attempt in Adolescent and Young Adult Twins

Gladys Langi

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>

 Part of the [Behavior and Behavior Mechanisms Commons](#), [Genetics Commons](#), and the [Other Psychiatry and Psychology Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/4757>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

The Influence of Genetic and Environmental Factors on Quit Attempt in Adolescent and Young
Adult Twins

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science
at Virginia Commonwealth University.

By
Gladys Emmanuella Putri Langi
BS (Institut Teknologi Bandung, July '12)

Director: Elizabeth C. Prom-Wormley, PhD, MPH
Assistant Professor
Division of Epidemiology

Virginia Commonwealth University
Richmond, Virginia
April, 2017

Acknowledgement

The author wishes to thank several people. First and foremost, I would like to thank my advisor and mentor Dr. Elizabeth C. Prom-Wormley. She has given me the utmost support and guidance during my studies at VCU and for completing this project. I would also like to thank my committee members, Dr. Hermine Maes and Dr. Michael Neale, for their time and thoughtful suggestions. I humbly thank Fulbright, the Institute of International Education, the American Indonesian Exchange Foundation, and the Department of Human and Molecular Genetics at VCU for their generosity and constant support for my studies in the USA. I also wish to thank my friends, Megan Hept, Dana Lapato, Javeria Aijaz, Anna Kovilakath, and Harini P. Tirumala, who have always been there for me. Finally, I would like to thank my partner, Nikodem Wojciechowski, and my parents for always loving and supporting me.

Table of Contents

Acknowledgement	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
List of Appendices	vii
List of Abbreviations	viii
Abstract	x
Introduction	11
Smoking Problem	11
Definition of cigarettes	11
Detrimental for population health	12
The complex process of developing and ending smoking behaviors	13
Steady decline in QA prevalence in adolescents	15
Difficulty in achieving smoking abstinence in adolescents	16
Limited understanding on the factors for SC in adolescents	17
The Genetic Epidemiology for Smoking Behaviors	17
The concept of family and twin studies	17
Genetic association studies	30
Epidemiology Studies: Environmental Factors	35
Summary	37
Statement of Study Aims and Hypotheses	38
Methods	40
Participants and Measures	40
Add Health: The National Longitudinal Study of Adolescent to Adult Health.....	40
MASATS: The Mid-Atlantic School Age Twin Study.....	41
VTSABD/TSA: Virginia Twin Study of Adolescent Behavioral Development/Transitions to Substance Abuse	41
Smoking Behavior Variables Used in The Study	42
Statistical Analyses	44
Testing age differences	44
Descriptive statistics	45

Twin modeling: CCC model for SI, SA, and QA	49
Model testing	53
Calculating proportion of variance components	54
Software	54
Results	56
Summary Statistics	56
Genetic Analysis	57
Testing sex differences	57
Testing causal pathways	58
Testing A and C and determining best fitting models	58
Path estimates for ACE models	61
Path estimates for best fitting models	62
Variance component estimates for QA from best fitting models	62
Discussion	75
First Aim: Estimating A, C, and E for QA	75
Second Aim: Testing sex differences for QA	78
Third Aim: The relationship between SI, SA, and QA	79
Sample differences	80
Limitations	81
Conclusion	81
Bibliography	84
Appendix	99
Vita	202

List of Tables

Table 1 Samples and Measures Used.....	44
Table 2 Data Groups Based on Age for This Study.....	45
Table 3 Prevalence of Smoking Behaviors of by Age and Sex for Add Health.....	46
Table 4 Prevalence of Smoking Behaviors of by Age and Sex for MASATS.....	46
Table 5 Prevalence of Smoking Behaviors of by Age and Sex for VTSABD/TSA	47
Table 6 Tetrachoric and Polychoric Correlations for Smoking Behaviors by Sample, Age, Sex, and Zygosity	48
Table 7 The 11 Models Testing The Significance of Pathways Estimates	50
Table 8 Equations for Calculating Proportion of Variances	54
Table 9 Model Comparison by Age Group: Add Health Subjects.....	58
Table 10 Model Comparison by Age Group: MASATS Subjects.....	60
Table 11 Model Comparison by Age Group: VTSABD/TSA Subjects.....	61
Table 12 Path Coefficients and 95% Confidence Intervals of ACE Models Without Quantitative Sex Differences for Each Data Group.....	64
Table 13 Variance Component Estimates from Best Fitting Models.....	74

List of Figures

Figure 1. Stages for The Development of Smoking Behaviors.	15
Figure 2. Uniphenotype ACE Twin Model.	19
Figure 3. Modeling Sex Differences in Twin Studies.	20
Figure 4. Single versus Multiple Liability Distributions for Smoking Behaviors (Maes et al. 2004).	22
Figure 5. Model 1: CCC ACE Model With Qualitative and Quantitative Sex Differences.	50
Figure 6. The Best Fitting Model and Path Estimates for Younger Adolescents (age 12-15): Add Health.	65
Figure 7. The Best Fitting Model and Path Estimates for Younger Adolescents (age 12-15): MASATS.	66
Figure 8. The Best Fitting Model and Path Estimates for Older Adolescents (age 16-19): Add Health.	67
Figure 9. The Best Fitting Model and Path Estimates for Older Adolescents (age 16-19): MASATS.	68
Figure 10. The Best Fitting Model and Path Estimates for Young Adults (age 22-32): Add Health.	69
Figure 11. The Best Fitting Model and Path Estimates for Young Adults (age 22-32): VTSABD/TSA.	70
Figure 12. The Best Fitting Model and Path Estimates for All Participants: Add Health.	71
Figure 13. The Best Fitting Model and Path Estimates for All Participants: MASATS.	72
Figure 14. The Best Fitting Model and Path Estimates for All participants: VTSABD/TSA.	73

List of Appendices

Appendix I R script for building datasets (ABD: VTSABD/TSA; ADH: Add Health; NCT: MASATS)	99
Appendix II R script for CCC Model for Add Health age 12-15. (ADH: Add Health).....	109
Appendix III R script for CCC Model for Add Health age 16-19. (ADH: Add Health).....	119
Appendix IV R script for CCC Model for Add Health age 22-32. (ADH: Add Health).....	130
Appendix V R script for CCC Model for Add Health all participants. (ADH: Add Health).....	141
Appendix VI R script for CCC Model for MASATS age 12-15. (NCT: MASATS)	151
Appendix VII R script for CCC Model for MASATS age 16-19. (NCT: MASATS).....	161
Appendix VIII R script for CCC Model for MASATS all participants. (NCT: MASATS).....	171
Appendix IX R script for CCC Model for VTSABD/TSA age 22-32. (ABD: VTSABD/TSA)	181
Appendix X R script for CCC Model for VTSABD/TSA all participants. (ABD: VTSABD/TSA)	192

List of Abbreviations

-2LL – -2 log-likelihood

Δ df – difference in degrees of freedom

Δ LL – difference in -2 log-likelihood

A – Additive genetic factors

ACE model – Twin model estimating A, C, and E

Add Health – National Longitudinal Study of Adolescent to Adult Health

AE model – Twin model estimating A and E

AIC – Akaike Information Criterion

C – Shared environmental factors

CCC – Common-Contingent-causal model

CE model – Twin model estimating C and E

CO – Carbon monoxide

CPD – Cigarettes per Day

D – Dominance

df – degrees of freedom

DZ – Dizygotic twins

DZF – Dizygotic females

DZM – Dizygotic males

DZO – Dizygotic opposite sex twins

E – Unique environmental factors

ep – number of estimated parameters

FDA – The Food and Drug Administration

GWAS – Genome-Wide Association Studies

H^2 – Broad sense heritability (H^2)

h^2 – Narrow sense heritability (h^2)

HR – Hazard Ratio

MASATS – Mid-Atlantic School Age Twin Study

MZ – Monozygotic twins

MZF – Monozygotic females

MZM – Monozygotic males

ND – Nicotine Dependence

p – p-Value

PAH – polycyclic aromatic hydrocarbons

QA – Quit Attempts

r –Correlation

re – Environmental correlations

rg – Benetic correlations

RS – Regular Smoking

SA – Smoking Amount

SB – Smoking behavior

SC – Smoking Cessation

SI – Smoking Initiation

SNP – Single Nucleotide Polymorphisms

VTSABD/TSA – Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up

WS – Withdrawal Symptoms

Abstract

THE INFLUENCE OF GENETIC AND ENVIRONMENTAL FACTORS ON QUIT ATTEMPT IN ADOLESCENT AND YOUNG ADULT TWINS

By Gladys Emmanuella Putri Langi, BSc

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2017

Major Director: Elizabeth C. Prom-Wormley, PhD, MPH, Assistant Professor, Division of Epidemiology

The purpose of this study was to examine the genetic and environmental influences on lifetime quit attempt in three US adolescent and young adult twin samples (N total = 6,322). The study utilized a common-contingent-causal (CCC) model to estimate these factors for lifetime quit attempts, after accounting for the factors for lifetime cigarette use and cigarettes per day. The study also examined age and sex differences, as well as the degree of relationship between these smoking phenotypes. The results demonstrated significant genetic influences for lifetime quit attempts in adolescents and young adults. No sex differences were observed for the contributions of genetic and environmental factors for lifetime quit attempts. Furthermore, separate liabilities for lifetime quit attempts and lifetime cigarette use were found for most age groups. Study findings have important implications for promoting quit attempts in adolescents and young adults.

Introduction

Smoking Problem

Definition of cigarettes

Cigarettes are rolled tobacco blend in paper intended to be burnt to release a smoke of chemicals. Manufacturers combine cured tobacco leaves with reconstituted tobacco, a combination of tobacco dust, stems and product reclaim, before adding numerous additives, such as ameliorants (sugars, cocoa, licorice) and enhancers of nicotine delivery (ammonium sulfide, diammonium hydrogen phosphate, and ammonium hydroxide) (Wigand 2006). In the US, 616 intentional additives are added to the tobacco blend (Wigand 2006). Manufacturers claim these additives are safe because they are listed as ‘Generally Recognized as Safe’ by the American Food and Drug Administration (FDA). However, while these ingredients are claimed to be safe when digested, it is unclear if they are safe when burnt and inhaled. It is most likely that these additives are harmful to lungs because the respiratory system lacks digestive enzymes to degrade the chemicals. In addition, these additives can convert to harmful substances when pyrolyzed. For example, sugars can convert to aldehydes that intensify the effect of nicotine. Glycerol converts to acrolein, a probable human carcinogen. Since industries are protected from fully disclosing their ingredients and pharmacological effects of all additives, smoking cigarettes poses a serious health risk to the public (Wigand 2006). Due to these additives, cigarette smoke contains a toxic mix of 7,000 chemicals (Rodgman and Perfetti 2013).

Detrimental for population health

Cigarette's toxic smoke immediately affects the human body. The delivery of cigarette smoke can vary greatly among smokers, depending on the cigarette design and smoking patterns, such as number of puffs and puff volume (US Department of Health and Human Services 2010). Nicotine and other chemicals are rapidly absorbed in the lung and are ultimately transported to other organs via the circulatory system (US Department of Health and Human Services 2014). Numerous animal and non-animal models have demonstrated the toxicity of cigarette smoke (US Department of Health and Human Services 2010). Cigarette smoke and its condensate have been demonstrated to be tumorigenic and mutagenic in various laboratory modelling systems (US Department of Health and Human Services 2010). For example, a recent human study using labeled polycyclic aromatic hydrocarbons (PAH), a likely major causative agent for lung cancer, demonstrated rapid formation of carcinogenic form of PAH in smokers (Zhong et al. 2011). This carcinogenic form reaches maximum levels within 15-30 minutes after smoking and reacts readily with DNA to initiate carcinogenic processes (Zhong et al. 2011). All together, these laboratory results support epidemiologic observations that smokers have worse health expectancies (US Department of Health and Human Services 2014) and shorter lifespan (Jha et al. 2013).

There are at least 16 million Americans suffering from a smoking-related disease (US Department of Health and Human Services 2014). In 2000, 8.6 million Americans had an estimated 12.7 million smoking-attributable serious medical conditions (Centers for Disease Control and Prevention 2003). Further, tobacco use causes 6 million deaths annually worldwide and it is estimated to cause 8 million deaths per year by 2030 (World Health Organization 2011). In the US, the annual smoking-attributable mortality for 2010-2014 is about 480,000 due to

smoking and secondhand smoke (US Department of Health and Human Services 2014). The estimated total of deaths for 2010-2014 is approximately 2.4 million deaths. This includes 818,500 deaths due to cancer, 803,000 deaths due to cardiovascular diseases, 565,500 deaths due to pulmonary diseases, and 206,420 deaths due to secondhand smoke exposure (US Department of Health and Human Services 2014). Thus, since the first Surgeon General's report on smoking and health in 1964 until 2014, there have been at least 20.4 million deaths due to smoking and exposure to secondhand smoke (US Department of Health and Human Services 2014). In the future, it is projected that there will be 5.6 million potential deaths, based on current youth smokers data across the US (US Department of Health and Human Services 2014).

The complex process of developing and ending smoking behaviors

The degree of smoking differs greatly among smokers. This is because there are several stages of smoking (Amos, Spitz, and Cinciripini 2010), as shown in Figure 1. There are two major processes: the stages for developing Nicotine Dependence (ND) (black ellipses) and the stages for Smoking Cessation (SC) (white ellipses).

Stages for ND development

The stages for ND development are Smoking Initiation (SI), Regular Smoking (RS), and ND. These stages have a unidirectional causality relationship with SI as the starting point. All smokers start with Smoking Initiation (SI) and experiment with smoking. Nearly all smokers start by age 18 (Elders et al. 1994; Khuder, Dayal, and Mutgi 1999; Wilkinson et al. 2007; Johnston et al. 2011). In this stage, smokers smoke intermittently or occasionally. Some smokers are able to maintain this low-level of smoking, but others continue to smoke daily and continuously. These smokers are considered to have entered the RS stage. In this stage, smokers can have differing level of smoking, where some smoke lightly (less than 10 cigarettes per day)

(Okuyemi et al. 2002; Husten 2009; Centers for Disease Control and Prevention 2011) and some smoke heavily (more than one pack per day) (Centers for Disease Control and Prevention 2011; Wilson et al. 1992). This continued use of tobacco products then drives the development of Nicotine Dependence (ND) or tobacco use disorder. ND is a medical condition characterized with symptoms of consuming large quantities of tobacco over a long period, developed tolerance for nicotine, and experience of withdrawal symptoms upon discontinued use of tobacco products (American Psychiatric Association 2013). In order to be diagnosed with ND, at least 2 of the 15 sub features need to be endorsed within a 12-month period (American Psychiatric Association 2013). The level of severity differs among smokers, where some have a mild dependency (2-3 symptoms), moderate dependency (4-5 symptoms), or severe dependency (6 or more symptoms) (American Psychiatric Association 2013). Thus, there is a great variety of smokers in developing ND.

Stages for SC process

To end their smoking habit, smokers undergo a process called Smoking Cessation (SC) (DiClemente et al. 1991). The stages for SC are Quit Attempts (QA), experiencing Withdrawal Symptoms (WS), responding to treatment, and Abstinence. There are bidirectional relationships between these stages. Once a smoker is ready to quit smoking, a QA is attempted by discontinuing tobacco use. Some smokers can remain free from tobacco use with only one attempt and successfully achieve the Abstinence stage (no tobacco use for a prolonged period, typically 6-12 months). However, a lot of smokers experience WS that can discourage them from giving up cigarettes. Failure to respond to treatment can also disrupt the SC process. For these smokers, it can take several QAs and trying numerous SC methods before finally achieving Abstinence.

Unclear relationship between ND development and SC process

In recent years, there have been growing evidence of the relationship between ND and cessation. ND is shown to be one of the major barriers for achieving Abstinence (Asfar et al. 2008; Gruder et al. 2012). Measurements of ND, such as the number of cigarettes consumed per day and ND diagnosis scores, can also predict Abstinence (Razavi et al. 1999; Breslau and Johnson 2000; Dale et al. 2001). However, it is less understood how the risk factors for ND development influence the initiation and success of SC process.

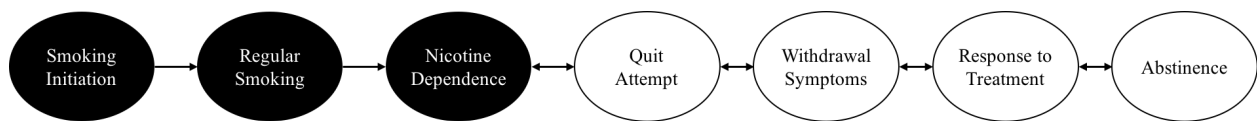


Figure 1. Stages for The Development of Smoking Behaviors. Black ellipses indicate smoking stages for ND development. White ellipses indicate smoking stages for SC process. The pathways illustrate the relationship between each stages.

Steady decline in QA prevalence in adolescents

Smoking abstinence is crucial for improving health quality. SC greatly alleviates the health consequences of smoking, for example, reducing the risk for bladder cancer, pancreatic cancer, and respiratory infections (U.S. Department of Health and Human Services 1990). Former smokers also reach rates of decline in pulmonary function comparable to never smokers (US Department of Health and Human Services 2004). Due to these health benefits, former smokers regain 6-10 years of life expectancy (Jha and Peto 2014). So, quitting smoking increases the life quality of smokers (U.S. Department of Health and Human Services 1990; Jha et al. 2013).

Despite well-known benefits of quitting smoking, the prevalence of smokers attempting to quit is decreasing in adolescents. Over a 10-year period from 2003 to 2013, the proportion of adolescent smokers attempting to quit decreased from 60.1% to 55.7% (Healthy People 2020 2016). This takes the US further away from its goal of 64% in 2020.

This is discouraging because adolescents are at high risk of developing ND (DiFranza et al. 2007) and smoking-related diseases. Adolescent light smokers are at high risk for smoking more cigarettes (Wiener, Shockey, and Morgan 2016) and for smoking regularly throughout adulthood (Elders et al. 1994). Most adolescent smokers are dependent on nicotine (Elders et al. 1994), even when having a low level of smoking (Goriounova and Mansvelder 2012). This early dependence on nicotine may disrupt the maturation process of the prefrontal cortex in adolescents, resulting in impaired cognitive ability, attention, and mental health (Goriounova and Mansvelder 2012). In addition, the greatest benefit of quitting smoking is when it is obtained when young (Panday et al. 2007). Some health benefits may only be achieved after years of non-smoking. For example, the reduction in risk for pancreatic cancer was measurable after 10 years of abstinence (US Department of Health and Human Services 1990).

Difficulty in achieving smoking abstinence in adolescents

One reason for this decline is the difficulty in achieving smoking abstinence in adolescents. Adolescents also experience withdrawal symptoms upon discontinuing cigarette use (Rojas et al. 1998; Prokhorov et al. 1996; Stanton, Lowe, and Gillespie 1996). Although ND smokers can benefit from SC programs that provide pharmacological aid and counseling, some of these methods are inefficient or unavailable for adolescents (Sims 2009; Karpinski, Timpe, and Lubsch 2010). No smoking cessation medications are approved by FDA for use in children

or adolescents (Karpinski, Timpe, and Lubsch 2010). Some guidelines approve the use of Nicotine Replacement Therapies, but other guidelines suggest case-by-case decisions (Karpinski, Timpe, and Lubsch 2010). Smoking cessation programs are usually developed for adults and these approaches have been shown to be unappealing and ineffective in younger smokers (Sussman 2003).

Limited understanding on the factors for SC in adolescents

Genetic Epidemiology studies have been done for estimating the contributions of genetic and environmental factors for smoking behaviors, including SC. Family and twin studies estimate the proportion of variance for SC that is accounted for by genetic and environmental factors. Genetic discovery studies and epidemiology studies further test which genetic and environmental factors are important for SC. Overall, there is evidence for genetic and environmental factors for smoking behaviors, but most studies are done in adults.

The Genetic Epidemiology for Smoking Behaviors

The concept of family and twin studies

Family studies

Family studies have reported that siblings of habitual smokers (smoke 1 pack daily) are 77% more likely to also become habitual smoker (Bierut et al. 1998). This implies that there is a shared liability of smoking between family members, which can be in form of shared genetic or environmental factors. However, family studies alone cannot differentiate these two factors separately. To properly estimate which factors are more important for this liability, a twin study is needed.

Twin studies

Twin studies are powerful in estimating the contributions of genetic and environmental factors on trait variation. Using twins and their extended family members, these studies can calculate how much of the variance of a trait is due to the additive genetic factors (A), dominance (D), shared environmental factors (C), and unique environmental factors (E). A and D encompasses the hereditary information passed on through generations for a particular trait, where A refers to additive effects of alleles at every locus (Cherny 2009). The proportion of trait variance that is due to A is also called the narrow sense heritability (h^2). Broad sense heritability (H^2) is the proportion of trait variance due to all genetic factors, including A, D, and epistatic interactions (Wray and Visscher 2008). C are environmental factors shared between twins, such as prenatal environment, social class, community, and schools (Plomin 2011). E are non-shared environmental factors that make twins different from each other. This E component also includes random error or measurement error from the study. In the classical twin study, only three factors can be estimated within one model (A, C, E or A, D, E). The model estimating A, C, and E (ACE model) usually fits smoking behavior phenotype data better than the ADE model (Madden et al. 2004).

Twin studies use monozygotic (MZ) twins and dizygotic (DZ) twins to study the relationship of these factors. MZ twins result from the fertilization of one egg and one sperm, so they have the same genetic material. DZ twins result from two different eggs and share, on average, 50% of their genetic material. Regardless of zygosity, twins have environmental factors which are shared completely between twins and environmental factors which are unique to each twin. These relationships for genetic and environmental factors between twins are illustrated in Figure 2. The trait correlation between MZ pairs and DZ pairs provide information on the

contributions of A, C, and E for that trait. For example, if MZ pairs are more similar for the trait than DZ pairs, then some genetic factors are expected to influence the trait. If DZ pairs are equally or more similar for the trait than MZ pairs, then the trait is expected to have substantial environmental influences. The components are estimated through maximum likelihood to find the combination of these factors that best explains the data (Neale 2009).

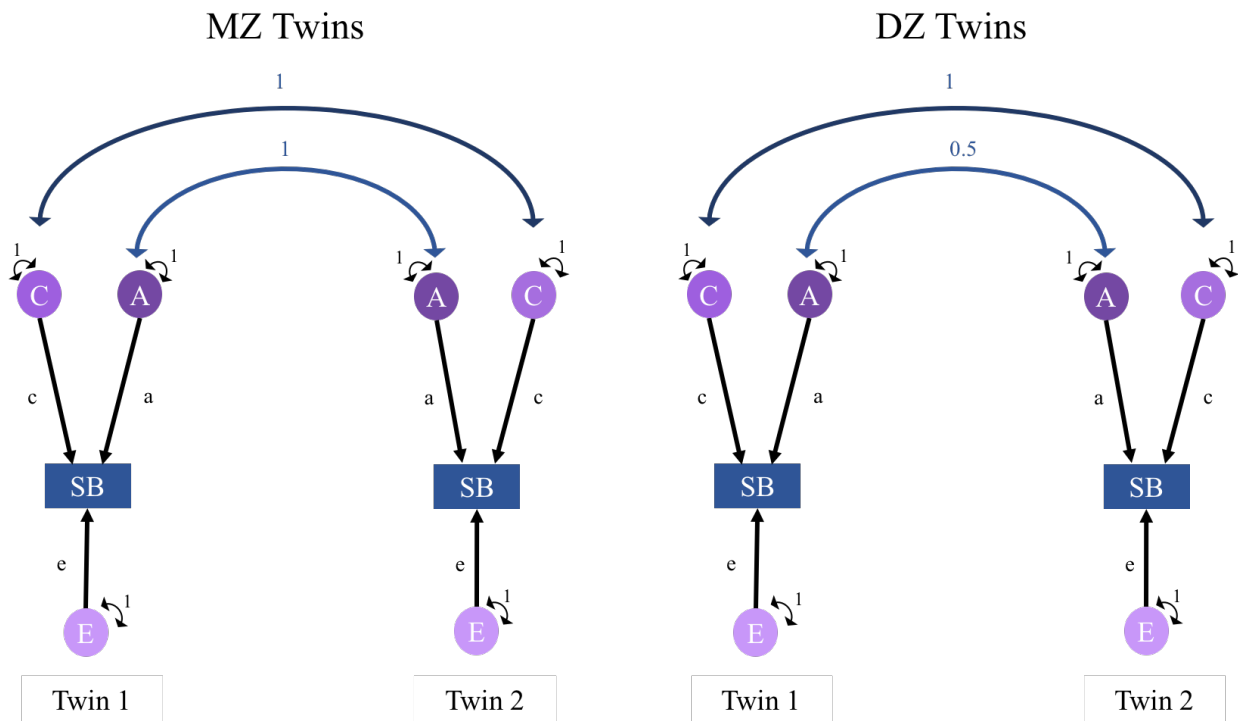


Figure 2. Uniphenoype ACE Twin Model. In this figure, squares illustrate observed smoking phenotype (SB: smoking behavior), circles illustrate latent factors A, C, and E. The latent factors A and C are correlated between the twins. The C factors are shared completely in both twin zygosity. The A factors are shared completely between MZ twins, but only shared half between DZ twins.

Twin studies can test the importance of sex differences in the contributions of A, C, and E. There are two types of sex differences: quantitative and qualitative sex differences. If quantitative sex differences exist, the genetic and environmental factors influence males and females differently. If qualitative sex differences exist, males and females have different sets of

A or C. To model quantitative sex differences, the a, c, and e pathways in twin models are simply estimated differently by sex for all zygosity groups (Figure 3). Opposite-sex pairs provide information about qualitative sex differences. The genetic/shared environmental correlations (rg/rc) between twin 1 and twin 2 are set to be estimated freely (Figure 3). An rg estimate of 0 indicates that the genetic factors for males are completely independent from those for females. In contrast, an rg of 1 means that the genetic factors for males are the same as for females (Neale and Cardon 1992).

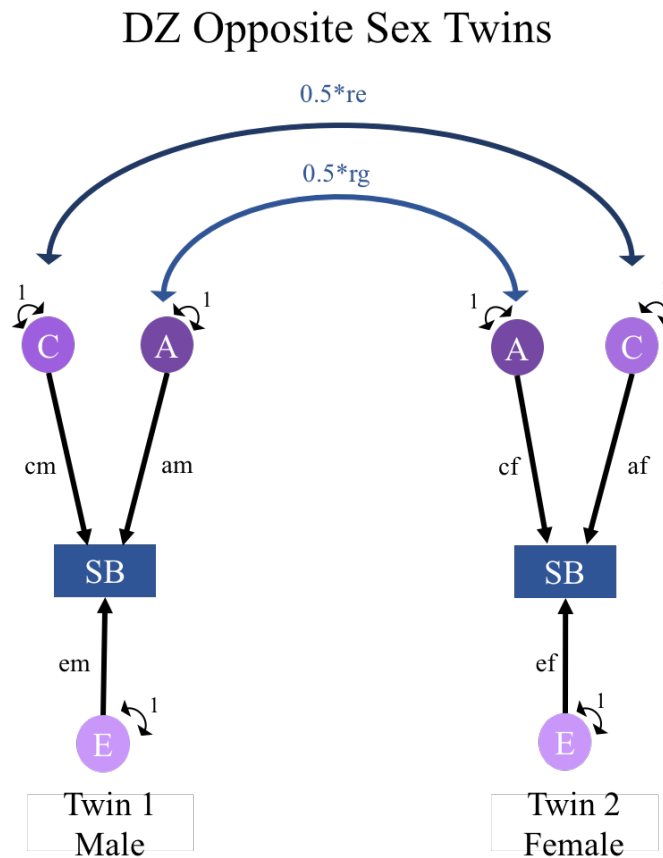


Figure 3. Modeling Sex Differences in Twin Studies. Quantitative and qualitative sex differences are pictured here for DZ opposite-sex pairs. Quantitative sex differences are modeled through a, c, and e pathway estimates for each sex (males: am , cm , em ; females: af , cf , ef). Qualitative sex differences are modeled through the genetic (rg) and shared environmental (re) correlations.

Multiphenotype twin studies

Since smoking behaviors are connected to each other (Figure 1), multiphenotype twin studies are used to model the relationship between smoking stages. These studies explore the nature of relationship of the liabilities for SI and its other smoking stages (RS, ND, QA, smoking abstinence). The two extremes are (1) the liabilities for SI and smoking stage are in a single continuum or (2) they are completely independent from each other (Figure 4). In the first scenario, the genetic and environmental factors for initiation also contribute to later stages, such as dependence (Heath et al. 2002). In the second scenario, the risk for becoming a smoker is uncorrelated from the risk of becoming dependent on nicotine. The addition of non-smokers to the analysis depends on the relationship nature. If the liabilities are independent to each other, inclusion of non-smokers as non-dependent individuals would confound two traits having different modes of inheritance (Heath et al. 2002). If the liabilities are correlated, then excluding non-smokers will discard genetic information and lead to biased genetic and environmental factor estimates for dependence (Heath et al. 2002). However, non-smokers have no data on dependence measures. So, genetic models need to be developed to test these relationships and to properly handle missing dependence data for non-smokers.

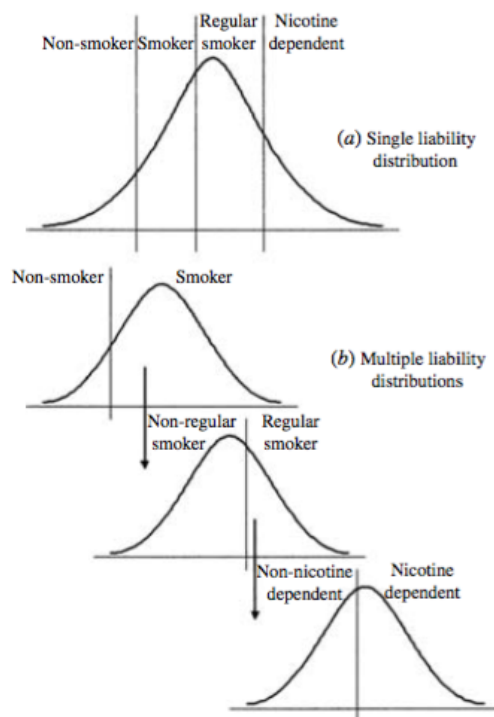


Figure 4. Single versus Multiple Liability Distributions for Smoking Behaviors (Maes et al. 2004). This is an example for the relationship between stages for ND progression. The liabilities for the stages lie in a single continuum in the single liability distribution model. If the single liability distribution model is rejected, then the liabilities for these stages can be independent or correlated (Maes et al. 2004).

There are two genetic models developed for such purposes: the common-contingent-causal (CCC) model (Kendler et al. 1999) and the bivariate genetic model with multicategory initiation variable (Heath et al. 2002). Since a non-user cannot also be a persistent or dependent user, a traditional bivariate model using a binary measure of initiation cannot be resolved to estimate genetic and environmental variances and correlations for initiation and outcome (Heath and Martin 1993). The CCC model overcomes this by applying a constraint on the model: genetic and environmental factors on initiation can only affect dependence through the observed phenotype of initiation, which is called a ‘common’ pathway (Kendler et al. 1999). ‘Contingent’ is used to describe the condition that dependence assessments can only be obtained from those who have initiated smoking. ‘Causal’ is for assuming a unidirectional path from smoking initiation to

outcome. This model then estimates the value of this causal path and the remaining genetic and environmental variance specific to a later stage. If the causal path is one, then there is a single liability relationship between initiation and progression. If the causal path is zero, then the liability for initiation and progression is orthogonal (Kendler et al. 1999). However, this CCC model assumes that shared environmental factors of smoking initiation must also have an effect on outcome in those who have smoked (Heath et al. 2002; Morley et al. 2007). This model performed poorly with some simulation samples, resulting in underestimated genetic correlation and biased parameter estimates (Heath et al. 2002). However, when only a binary initiation measure is available, the CCC model can still be used but a sensitivity analysis is needed to explore point parameter estimates and their 95% confidence intervals (Heath et al. 2002).

The second model is a bivariate model with multicategory initiation phenotype proposed by Heath et al. In this model, age of initiation is utilized to categorize SI into three categories: never smokers, early smoker (those initiating before 18 years old), and later onset smoker (initiating regular smoking after 18 years old). This model is preferred over the CCC model as it has been shown to be more powerful in estimating correlation between initiation and other smoking stages (Heath et al. 2002). However, having a multicategory initiation is not always possible as not all smoking samples have asked age of SI.

These models can estimate the degree of liability overlap between the stages. The bivariate model with multicategory initiation phenotype model can estimate the correlations of the genetic and environmental factors between stages. The CCC model can estimate how much of the total proportion of variance is specific to an outcome stage and how much is shared with previous stages. Knowing this can help refine public health efforts. For example, if there are significant genetic influences specific for QA after accounting for SI and ND, then genetic

association studies for QA are worth pursuing. A substantial amount of shared genetic influence between smoking abstinence and previous stages would also indicate the need for accounting for genetic variants associated with SI and ND when analyzing genetic variants for QA. Failure to do so might result in inaccurate associations between genetic information and QA. Additionally, if a substantial amount of liability for QA is not shared with SI nor ND, then designing stage specific public health interventions is needed.

Twin studies findings for smoking initiation/regular use

Numerous twin studies have demonstrated the importance of A for SI with estimations ranging from 22% to 72% (Rose et al. 2009). Substantial contributions for C and E are also estimated for SI (C = 0-56%; E = 5-52%) (Rose et al. 2009). Most twin studies use ever/never regular smoking and some use ever/never cigarette use as the definition of SI. Most studies use adult twins as participants, while twin studies on adolescents have been done recently. The wide range of heritability estimates are likely due to age, sex, and population or study differences.

There are consistent findings that heritability estimates differ by age groups. Madden et al. demonstrated that A can be equated across three age groups of adult twins: young adults (18-25 years of age), adults (26-35 years of age), and older adults (36-45 years of age). However, C estimates differed between age groups, with young adults having a higher estimate (C = 33%-45%) than other age groups (C = 9-35% in adults, C = 11%-26% in older adults) (Madden et al. 2004). A different study in the US found age effects on heritability where older males have greater A influence (A = 79%) than younger males (A = 66%) (Hamilton et al. 2006). However, the same study found that estimates for females could be equated across age groups (Hamilton et al. 2006). Similar trends were found in a mega analyses of 11 adolescent twin studies (Maes et al. 2016), where A had increasing importance in SI, from 15% in age 13 to 45% in age 19. C had

a decreasing importance with increasing age (from 70% in age 13 to 40% in age 19), while E had a stable minimum influence in smoking initiation (Maes et al. 2016).

Sex differences on heritability estimates are somewhat conflicting. In a meta-analysis of six twin studies from the US, Australia, and Finland in the 1990s (Li et al. 2003), A were found to be important for SI, but the influence was stronger in females (females: A = 55%; males: A = 37%). Similarly, Heath et al. demonstrated a stronger genetic influence of SI in adult females (females: A = 63%; males: A = 22%). However, Hamilton et al. demonstrated an opposite relationship where A were more important in males (A = 71%) than in females (A = 32%). Madden et al. found that estimates for A were slightly higher in males (females: A = 46%; males: A = 57%), while the estimates for C were higher in females (females: C = 26%-45%, males: 9%-33%). In adolescents, sex differences in heritability estimates seem to be age dependent. The mega analyses of adolescent twins demonstrated that the correlation between shared environmental factors between males and females decreased from 1.0 in age 12 to 0.6 in age 19, suggesting differences between males and females are more pronounced in later age (Maes et al. 2016).

Twin studies also offer some insight on the generalizability of estimates for different countries/populations. A study of three samples from Sweden, Finland, and Australia demonstrated that A can be equated across countries (Madden et al. 2004). However, in the same study, C estimates can only be set equal across countries in females. In males, estimates for C can be equated between Sweden and Finland, but not with Australia (Madden et al. 2004). In the mega analysis of 11 twin studies of European descent, the prevalences of SI could not be equated across studies (Maes et al. 2016). This reflects differences in SI assessment between studies and culture differences concerning attitudes toward smoking (Maes et al. 2016). However, twin

correlations could be equated across studies, suggesting similar etiology of SI liability between these countries with European ancestry. Thus, preventive measures in one country of this study is likely to be effective in another (Maes et al. 2016).

A multiphenotype study has been done between SI and RS. In the study, SI was defined as ever/never cigarette use and RS as smoking on average of at least seven cigarettes per week for a minimum of 4 weeks (Maes et al. 2004). This study found that most of the genetic factors for RS were shared with smoking initiation and 21% were specific to RS (Maes et al. 2004). The causal pathway from SI to RS was high ($b = 0.68-0.89$), suggesting highly overlapping liabilities between these phenotypes.

Twin and family studies for nicotine dependence

Some studies use ND diagnosis as the phenotype of interest, some use smoking amount (SA) measured as cigarettes per day (CPD). CPD has been demonstrated to correlate with ND diagnosis measures (Heatherton et al. 1989; Lichtenstein and Mermelstein 1986; Piper et al. 2004; Shiffman and Sayette 2005). CPD is often easier to obtain in studies and its high correlation with ND makes it a good proxy for ND.

Additive genetic factors and unique environmental factors have substantial influence on ND. When measured using CPD, A and E accounted for most of the total variance, but some C was reported ($A = 40\%-61\%$, $C = 0\%-30\%$, $E = 18\%-51\%$) (Rose et al. 2009). For ND diagnosis, twin studies have consistently reported substantial A influence with no influence from C ($A = 30\%-75\%$, $C = 0\%$, $E = 25\%-70\%$), even with various diagnosis measures and in different cultures (Rose et al. 2009; Broms et al. 2007).

Similarly to SI, there are conflicting findings for sex differences in the etiology for ND. A meta analysis of adult twins by Li et al. (2003) found that A and E were important in both genders, but slightly more in males (males: A = 59%, E = 37%; females: A = 46%, E = 24%). C were virtually none in males, but moderately important in females (males: C = 8%; females: C = 28%) (Li et al. 2003). Broms et al. (2006) demonstrated different sex effect for SA, where heritability estimates were higher in women (males: A = 54%; women: A = 61%). However, this study did not formally test sex differences as opposite-sex twin pairs were not included in the analysis (Broms et al. 2006). Most studies with opposite-sex twin pairs found no sex differences for ND (Vink et al. 2004; Maes et al. 2004; Heath et al. 2002; Madden et al. 2004; Hamilton et al. 2006).

Multiphenotype studies have studied the relationship between SI and SA/ND. SI was shown to have a substantial genetic correlation with SA (males: $r_g = 0.84$, females: $r_g = 0.59$) (Morley et al. 2007), but another study found a weak genetic correlation in males ($r_g = -0.22$) and females ($r_g = -0.17$) (Broms et al. 2006). Using the CCC model, studies estimated that 24-31% of the total genetic factors for ND was specific for this stage, after accounting for SI (Kendler et al. 1999) and RS (Maes et al. 2004). Causal pathways estimated was negative from SI to ND, measured as Fagerström Test for Nicotine Dependence (FTND) score, ($b = -0.29$) in a study (Maes et al. 2004), but high in a study of female twins ($b = 0.77$), where ND was measured using the Fagerström Tolerance Questionnaire and DSM-IV (Kendler et al. 1999). However, when a dichotomized measure of ND was used, the causal path from SI to ND was no longer significant (Maes et al. 2004). Between RS and ND, a high estimate for the causal pathway ($b = 0.69-0.93$) was demonstrated (Maes et al. 2004).

However, recent multiphenotype studies for adolescents demonstrated a different relationship between SI and SA. In adolescents aged 14-15, SI and SA had independent liabilities (Bares, Kendler, and Maes 2015; Do et al. 2015). The causal pathway was not significant and could be dropped from the model (Bares, Kendler, and Maes 2015; Do et al. 2015). The liabilities for SI and SA were correlated from age of 16 onwards (Bares, Kendler, and Maes 2015) (age 16-17: $b = 0.93$; age 18-25: $b = 0.62$; age 26-33: $b = 0.65$). With this high overlap with SI, specific A and C for smoking quantity were no longer significant (Bares, Kendler, and Maes 2015). However, another study demonstrated independent liabilities for age group 16-17 and a significant causal pathway in group age groups (age 22-32) (Do et al. 2015). Additionally, unlike the participants in Bares et al., this study demonstrated a significant C for SA in the adolescent groups (Do et al. 2015). Eventhough both studies used individuals from the United States, the Bares et al. study used a nationally representative sample, while Do et al. used twin samples collected from Virginia. State policies and social acceptance towards smoking behaviors might influence the differences between these studies's smoking prevalence and parameter estimates. It is interesting to explore how generalizable adolescent genetic and environmental estimates are among different populations.

Twin and family studies for smoking cessation

In cessation studies, the binary measure of smoking abstinence (former/current smoker) is typically used. The criteria for smoking abstinence vary between studies. Most studies used six months as the cutoff (Lee and Kahende 2007; Hymowitz et al. 1997; Borland et al. 1991; Farkas et al. 1999), while one study considered three months of abstinence as successful (Hatziaandreu et al. 1990). However, some studies use more conservative criteria of one-year sustained abstinence (Xian et al. 2003; Zhu et al. 2000) and biomarker confirmation (Senore et al. 1998; Tonnesen et

al. 1999; Stapleton et al. 2006), especially in studies testing cessation treatments. Choosing a higher threshold for abstinence is preferable because 65-75% of those with less than 6 months of abstinence could relapse within a year (Bjornson et al. 1995; Murray et al. 2000). Some can also relapse after more than 2 years of abstinence (Murray et al. 2000).

A and E factors, but not C, were important for smoking abstinence (former vs current smoker), after accounting for SI (Heath and Martin 1993; Heath, Madden, and Martin 1998; Heath et al. 2002; Madden et al. 1999; True et al. 1997; Morley et al. 2007; Hamilton et al. 2006; Broms et al. 2006). Similarly, additive genetic and unique environmental factors significantly influenced failed QA and WS in adult male twins (Xian et al. 2003). Failed cessation was defined as having made a QA and reporting smoking within 12 months (Xian et al. 2003). Using a bivariate model, this study demonstrated moderate correlation between the A factors for failed QA and WS ($r_g = 0.31$) and a small environmental correlation ($r_e = 0.16$). A effects accounted for 54% of the total variance for failed QA and 29.7% for WS (Xian et al. 2003).

Sex differences were found for successful quitting in adults. Although the estimate value was similar between males and females (males: A = 58%, E = 43%; females: A = 50%, E = 50%), the parameters cannot be equated between sexes (Broms et al. 2006), indicating sex differences in the genetic and environmental factors for becoming a former smoker. Additionally, a small genetic correlation existed between SI and smoking abstinence in males ($r = 0.22$), but not in females (Broms et al. 2006). In females, there was a negative correlation of environmental factors between SI and smoking abstinence ($r = -.013$) (Broms et al. 2006).

To our knowledge, there are no twin studies for lifetime QA (ever/never attempted a QA) or any other stages of SC in adolescents. Adolescents is a crucial stage where smoking should be prevented and terminated as early as possible. Our understanding on the influence of genetic and

environmental factors for promoting QA and maintaining smoking abstinence in adolescents is lacking. Thus, studies should be conducted to provide useful information for designing appropriate public health efforts for adolescent smokers.

Twin studies summary

Twin studies have demonstrated that A, C and E factors play a role in smoking behaviors. The estimates vary, sometimes significantly, between age groups, sex, population, and phenotype definition. Nevertheless, genetic factors have been shown to be important for smoking behaviors with increasing importance from SI until smoking abstinence. The C factors appeared to be more important for the early stages of smoking behaviors, while E had a significant importance in later stages. Twins studies provide the basis of further studies to specifically identify which genetic and environmental factors are important for each smoking stage.

Genetic association studies

There are several types of genetic association studies. Candidate gene studies analyze only a certain genetic region. The region selection is based on previous research findings or an established theory, such as from model organism studies. Genome-wide association studies (GWAS) are the opposite, in which genetic associations are studied agnostically. No prior knowledge or evidence is needed, as GWAS scans the entire genome using a set of genetic variant markers. These markers are common single nucleotide polymorphisms (SNPs) selected based on linkage disequilibrium patterns in certain populations (Bush and Moore 2012). Since GWAS markers only account for a small proportion of the total genome, information on the ungenotyped SNPs is obtained through imputation. Imputation is done using genomic databases, such as the HapMap Consortium database (Frazer et al. 2007; Li et al. 2009). To directly sequence the whole genome, next-generation sequencing is utilized (Behjati and Tarpey 2013).

Next-generation sequencing can sequence exomes only (a region in the genome that encodes proteins) or the entire genome. Exome sequencing is more cost-effective and more useful clinically than whole-genome sequencing. Whole-genome sequencing is particularly useful for analyzing rare variants, but a combination of exome sequencing and GWAS data can improve imputation accuracy of rare variants (Kim et al. 2015).

GWAS is the most common type of genetic study for smoking behaviors, as it is the most cost-effective method compared to other study designs. Since studies on smoking behaviors mostly use European descendants and genetic databases for this population are largely established, GWAS serves to be the most efficient way to study smoking behaviors in this population. The following sections will highlight findings from GWAS on each smoking stage.

Genetic association studies for smoking initiation/regular smoking

Genetic association studies have identified genetic variants associated with SI in adults. A meta-analysis of several GWAS discovered several single nucleotide polymorphisms (SNPs) in *BDNF* to be associated with SI, which was defined as ‘ever regularly smoked’ (Tobacco and Genetics Consortium 2010). The most significant SNP in *BDNF* (rs6265) was replicated in a recent GWAS of 2,447 participants (Agrawal et al. 2015), with a SI definition of smoking 100 or more cigarettes during the lifetime. This phenotype is also referred to as RS, but used as SI in prior meta-analyses (Agrawal et al. 2015). An earlier GWAS used a different question for SI (do/did you smoke cigarettes?) which includes regular and experimental smokers. The study found significant associations with nine SNPs upstream of *IL15*, the strongest association was with rs4956302 (Liu et al. 2009). However, when the authors use a different samples with a different question for smoking initiation (did you smoke cigarettes regularly in the last year?), the rs4956302 was no longer significantly associated. While factors such as population and

sample size difference might play a role, this suggests that different genetic variants might be important for different levels of smoking (experimental versus regular smokers). Further studies are needed to confirm this.

For adolescents, a genetic association study for SI and RS was done using a genetic risk score. This genetic risk score was developed from SNPs associated with SA in adults. The study demonstrated no association between the risk score with SI. However, there was an association between the risk score with RS where those with high genetic risk scores were more likely to become daily smokers and persisted longer in smoking heavily (Belsky et al. 2013). Further studies are needed to determine genetic variants for SI and RS specific for adolescents and whether there are developmental changes in the influence of these genetic variants.

Genetic association studies for nicotine dependence

Genetic studies have identified genetic loci associated with SA. Three loci were identified to be associated with this phenotype: *CHRNA3*, *LOC100188947*, and *EGLN2*, which is near *CYP2A6* (Tobacco and Genetics Consortium 2010). The synonymous SNP rs1051730 in *CHRNA3* showed the strongest association (Tobacco and Genetics Consortium 2010). The A allele of this SNP corresponds to increased cigarettes per day (beta = 1.03) (Tobacco and Genetics Consortium 2010). Another SNP, rs16969968, was the second most significant variant to be associated with cigarettes per day (Tobacco and Genetics Consortium 2010) and has been proposed as a causal variant (Saccone et al. 2009). The effect size for this SNP is similar (beta = 1.00), where each A allele corresponds to an increase in cigarettes per day (Tobacco and Genetics Consortium 2010).

Several studies also analyzed the association between genetic variants and ND. A study demonstrated 35 SNPs to be significantly associated with ND (Bierut et al. 2006). These SNPs

are located in several genes, including *CHRNA3*. The *CHRNA3* gene was also reported in a separate study to be strongly associated with ND, along with SNPs in the *CHRNA5* and *CHRNA3* genes (Saccone et al. 2006). A study by Uhl et al. in 2007 moderately supported associations between ND and the Cholinergic Receptor Nicotinic genes. In addition, it listed 32 genes that are called substance dependence vulnerability genes, because they were found in ND smokers and in substance abusers from other studies (Uhl et al. 2007).

Recent GWAS studies demonstrate stronger associations with smoking biomarkers, such as carbon monoxide (CO) and cotinine. CO levels are easier to obtain, but CO is a less stable measure than cotinine and may be a poor indicator of exposure in light or occasional smokers (Bloom et al. 2014). Cotinine has a longer half-life, but its levels are also dependent on metabolisms unrelated to smoking exposure, such as variation in nicotine and cotinine metabolism (Bloom et al. 2014). Nonetheless, the genetic variants associated with cigarettes per day have stronger associations with these biomarkers. The variant rs16969968 in *CHRNA5* was associated with CO levels more than SA and ND (beta for CO = 2.66; SA = 0.08; ND = 0.17) and explains a larger portion of CO variance (2.06%) than SA (0.47%) in European decents (Bloom et al. 2014). Similarly, the *CHRNA5-CHRNA3-CHRNA4* cluster had a higher effect size for cotinine (0.30) than SA (0.13) and explained a higher proportion of the variance in cotinine (4.3% versus 0.9% of SA) in European decents (Keskitalo et al. 2009).

A genetic association study identified 16 SNPs for ND in adolescents (O'Loughlin et al. 2014). This prospective study used five ND phenotypes derived from various ND measures and 321 SNPs in 24 candidate genes associated with SA. A total of 16 SNPs in *ANKK1*, *CHRNA7*, *DDC*, *DRD2*, *COMT*, *OPRM1*, and *SLC6A3* were associated with at least one phenotype. Most of these genes are involved in dopaminergic pathways (*ANKK1/DRD2*, *DDC*, *COMT*, *OPRM1*,

and *SLC6A3*) (O’Loughlin et al. 2014). This study was done in Canadian adolescents, so additional studies to confirm these findings in the US are needed.

Genetic association studies for smoking cessation

Several genes have been found to be associated with smoking abstinence in adults (Uhl et al. 2008). The tobacco consortium study identified rs3025343 near *DBH* to be associated with smoking abstinence (Tobacco and Genetics Consortium 2010). Another GWAS study analyzed smokers enrolled in three different clinical trials. This study found eight genes showing significant association with cessation success in all three samples and 55 genes with at least two nominally significant SNPs in two samples (Uhl et al. 2008). Other studies have also studied smoking abstinence, but in a candidate gene-based approach. To date, there are 23 candidate gene-based association studies identifying additional 12 separate genes (Wang and Li 2009). Some of these genes are shared with previous studies and work in common biological pathways and some are specific for smoking abstinence. For example, some genes in the amyotrophic lateral sclerosis signaling, circadian rhythm signaling, glutamate receptor signaling, and synaptic long-term depression are specifically associated with smoking abstinence and not other stages (Wang and Li 2009).

Some genetic association studies have been done for SC stages in adolescents. A candidate gene study demonstrated significant associations between SNPs in *CHRNA3* and the number of quit attempts in adolescents (Hoft et al. 2008). Another study found significant association between SNPs in the *OPRM1* gene with nicotine withdrawal sensitivity score (Hardin et al. 2009). However, there are no genetic association studies yet for lifetime quit attempts, cessation treatment response, and smoking abstinence in adolescents.

Summary of genetic epidemiology studies of smoking behaviors

There are numerous genes significantly associated with smoking behaviors. These findings support the importance of genetic factors on smoking behaviors. A study by Wang et al. found significant overlap between these reported genes. Five genes were included in all smoking stages: *COMT*, *CYP2A6*, *DRD2*, *DRD4*, and *SLC6A3*. Six genes were associated only with SI/RS and ND: *CHRNA3*, *CHRNA5*, *CHRNA3*, *PTEN*, *SLC6A4*, and *TPH1*. Ten other genes were associated with ND and smoking abstinence: *A2BP1*, *ARRB2*, *CDH13*, *CHRNA2*, *CSMD1*, *CYP2B6*, *DBH*, *OPRM1*, *PRKG1*, and *PTPRD*. These genes operate in numerous pathways and there are four common pathways for all stages: calcium signaling, cAMP-mediated signaling, dopamine receptor signaling, and G-protein-coupled receptor signaling (Wang and Li 2009).

However, these genetic studies have important limitations. Some studies have small sample sizes, heterogeneity, and no correction for multiple testing (Wang and Li 2009). This can result in false positive results in the genetic associations, especially for SI and QA. These phenotypes have not been studied as extensively as ND, so they have limited replication studies to confirm the findings. Additionally, most studies for SC focused on smoking abstinence and were conducted in adults. The genetic variants influencing adolescent smoking behaviors and lifetime QA might be different from adults.

Epidemiology Studies: Environmental Factors

Environmental risk factors for smoking behaviors have been studied through epidemiological research. Some of these risk factors make twin's smoking behavior similar with their co-twin, which are estimated as C in twin studies. C factors are significant for SI, but have little to no influence in other stages. Family-related environmental factors, for example parental

smoking behavior (Scherrer et al. 2012; Tyas and Pederson 1998; Turner, Mermelstein, and Flay 2004), parental attitudes toward their children's smoking behavior (Tyas and Pederson 1998; Farkas et al. 1999; Turner, Mermelstein, and Flay 2004; Emory et al. 2010), and socioeconomic status (Lund 2015; Turner, Mermelstein, and Flay 2004; Borland et al. 1991; Broms et al. 2004; Lee and Kahende 2007; Hymowitz et al. 1997) have been demonstrated to be associated with SI and smoking abstinence. Cultural attitudes and policies towards smoking behaviors also contribute to population-specific smoking prevalence. This includes bans on availability of cigarettes, cigarette price increases, and cigarette advertisements (Joossens 2006; Brown, Platt, and Amos 2014; Lund 2015; Tyas and Pederson 1998; Turner, Mermelstein, and Flay 2004).

In the later stages, E factors are more important than C. As twins grow older, they might have separate lives and are exposed to different environmental risk factors. These factors make twins develop different smoking status and behaviors. For example, working in different jobs with different smoking policies can impact a twin's decision to quit smoking. There is evidence that stricter no-smoking policies increase quit attempts (Borland et al. 1991; Lee and Kahende 2007). Peer smoking is also associated with SI and reduced QA success, so twins spending time with different friends can develop different smoking behaviors as well (Richmond, Austin, and Webster 1988; Tyas and Pederson 1998; Turner, Mermelstein, and Flay 2004; Park, Weaver, and Romer 2010). In addition, personal stressful life events might drive a twin to smoke more than their co-twin (Tyas and Pederson 1998). However, a careful interpretation of E should be done as this estimate also includes measurement error.

Summary

Smoking behavior is a complex trait that has been studied extensively in adults. Advanced understanding of the roles of genetic and environmental factors for SI, RS, and ND has been established. Twin studies have demonstrated the importance of genetic factors for smoking behavior stages. Gene discovery studies confirm these estimates by identifying important genetic variants for each smoking stage. Environmental factors also have a significant role, where C is significant for SI, while E is more important in later smoking stages. Epidemiological studies have identified these environmental risk factors associated with smoking behaviors.

Twin studies have demonstrated shared liability between smoking stages. For example, ample evidence is present for shared liability between SI and ND. Genetic and environmental factors for SI was partially shared with ND. For smoking abstinence, a small genetic correlation with SI has been demonstrated in adult male smokers. A moderate genetic correlation was also demonstrated between failed QA and WS in adult male twins.

However, adolescent smoking behavior is less studied. Current literature demonstrated significantly strong influence of C for SI and low contributions of this factor for SA in young adolescents. Recently, novel findings of the relationships between SI and SA demonstrated different parameter estimates for adolescents in contrast to adults. However, the results are quite conflicting and no study has been conducted so far to study lifetime QA in adolescents. Further studies to understand the underlying relationship of the genetic and environmental factors for smoking is crucial. This information is needed to guide gene discovery studies and public health policy decisions for adolescent smoking prevention and cessation.

Statement of Study Aims and Hypotheses

This study aims to explore QA and its relationship with SI and SA in adolescent and young adult twins living in the USA. QA is defined as lifetime QA, SI as lifetime cigarette smoking, and SA is measured as maximum CPD. The three smoking stages were fitted in a series of CCC models to estimate the contributions of genetic and environmental factors shared and specific to each stage. Three samples of US twins collected in the 1990s were used in the analysis: the National Longitudinal Study of Adolescent to Adult Health (Add Health), and the Mid-Atlantic School Age Twin Study (MASATS), and the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up (VTSABD/TSA).

The first aim was to estimate the genetic and environmental contributions for lifetime QA after accounting for SI and SA, by age group. Three groups are studied: younger adolescents (12-15 years of age), older adolescents (16-19 years of age), and young adults (22-32). We hypothesize that A and E factors would be important for lifetime quit attempt, but not C, akin to previous reports for successful cessation in adults (Heath and Martin 1993; Heath, Madden, and Martin 1998; Heath et al. 2002; Madden et al. 1999; True et al. 1997; Morley et al. 2007; Hamilton et al. 2006; Broms et al. 2006). We also hypothesize that there will be developmental differences for the genetic and environmental estimates for QA.

The second aim was to test sex differences in the nature and magnitude of genetic and environmental contributions for QA. We hypothesize that sex differences for the estimates for smoking cessation are significant, according to Broms et al. This aim will determine whether gender-focused public health policies are needed for adolescent smoking.

The third aim of this study was to explore the relationship between SI, SA, and QA by age group. The hypotheses are as follows: (1) As reported in previous studies (Bares, Kendler, and Maes 2015; Do et al. 2015), we expect an absence of a causal pathway from SI to SA in younger adolescents and a presence of this pathway in the older adolescents and young adults group. (2) We hypothesized that the causal path from SI to QA will be similar, where it is present in older adolescents and young adults, but absent in younger adolescents. This aim will help understand whether public health policies for QA can be shared with SI or not.

Methods

Participants and Measures

Add Health: The National Longitudinal Study of Adolescent to Adult Health

Add Health is a Nationally representative longitudinal study of American adolescents in grades 7-12. This study investigated the health and behaviors of adolescents and their outcomes in young adulthood. The study included 12,105 adolescents collected from 132 Americans in 1994-1995. Among the participants, 784 twins born were included. Subsequent waves were conducted at 1995-1996, 2001-2002, and 2008-2009. In the first wave, participants ranged in age from 12 to 17, while in the last wave, they aged from 28 to 33 years old. Data were collected from each twin, parents, school administrators, friends, and romantic partners. Zygosity was asked through self-report questions about twin's similarity and confirmed by matching on 12 molecular genetic markers (Harris et al. 2006).

Participants with missing zygosity, sex, and lifetime smoking initiation were excluded from the study. As a result, a total of 118 unique participants were excluded from Add Health. A total of 24 non-smokers were then recoded into smokers because they had a value of one or more for SA and have tried to quit smoking. After recoding, all remaining non-smokers were set to have missing values for smoking progression phenotypes. In total, there were 1,444 participants of Add Health, with 281 MZ males, 289 MZ females, 255 DZ males, 222 DZ females, and 397 DZ opposite sex individuals.

MASATS: The Mid-Atlantic School Age Twin Study

The MASATS is a cross-sectional study of adolescent twins living in North Carolina and Virginia (D'Onofrio et al. 1999). MASATS investigated the causes of adolescent behaviors, particularly substance abuse behaviors. 5,000 twins registered in the Mid-Atlantic Twin Registry were invited to participate. Questions of zygosity, which have been shown to correctly determine zygosity in 90% of twins, were also asked. For smoking behaviors, twins were asked about their lifetime cigarette use, number of cigarettes per day, and number of quit attempts.

There were some participants with missing zygosity, sex, and lifetime SI. These 142 participants were excluded from the study. There were 8 participants with conflicting smoking status, so they were recoded as smokers. Then, all non-smokers were recoded as missing for smoking progression phenotypes. In total, there were 2,081 participants of MASATS, with 336 MZ males, 561 MZ females, 281 DZ males, 395 DZ females, and 508 DZ opposite sex individuals.

VTSABD/TSA: Virginia Twin Study of Adolescent Behavioral Development/Transitions to Substance Abuse

VTSABD is a longitudinal cohort study conducted in Virginia of adolescent twins and their parents (Hewitt et al. 1997; Simonoff 1997). The purpose of this study was to elucidate the influence of genetic and environmental factors for child and adolescent behavioural problems. Data were collected in four waves from 1990 to 2007. The study included 1412 Caucasian families of twins born between 1974 and 1983. Data were collected from each twin, biological parents, teachers, and other parenting figures. Zygosity status was determined through parent's questionnaire. Subsequent waves were conducted at one-and-a-half-year intervals. Two young-

adult follow-ups was conducted after wave 4: Young Adult Follow-Up (YAFU) and Transitions to Substance Abuse (TSA). After agreeing to join TSA, the twins of VTSABD answered a new set of questions focusing on substance abuse risk factors. The age range of VTSABD participants was 8 to 18 and for TSA was 22 to 32.

A total of 25 participants were excluded due to missing zygosity, sex, and lifetime SI. One non-smoker was recoded as smoker because they indicated smoking 1 CPD. Other non-smokers were recoded to have missing values for smoking progression phenotypes. In total, there were 2,797 participants from VTSABD/TSA, with 630 MZ males, 840 MZ females, 362 DZ males, 376 DZ females, and 589 DZ opposite sex individuals.

Smoking Behavior Variables Used in The Study

A total of three observed variables for SI, ND/SA, and QA were chosen for this study. The questions for each measurement in each study are summarized in Table 1. SI was measured as a binary variable for lifetime cigarette initiation. Those who have tried a cigarette were coded as 1, while never-smokers were coded as 0. This measure was asked in all samples. Information about age of initiation was not available in Add Health, so grouping smokers into three groups as in Heath et al., 2002, was not possible. For multiwave samples, which are VTSABD and Add Health, the highest value of this variable across all waves was used for each participant. Meaning, if a participant answered “yes” (coded as 1) at wave 2, but answered “no” (coded as 0) at wave 3, their lifetime smoking use will still be 1.

SA was used as a proxy for ND because ND diagnosis was only available in the TSA wave of VTSABD. SA was measured as CPD. For multiwave samples, the maximum CPD value was used for each twin. Although each sample had multiple levels of cigarette amounts the twins

could choose from, ranging from 0 - 95 cigarettes per day, the number of twins endorsing each level was not enough for analysis. Therefore, in this study, smokers were grouped into non-daily smokers (0 CPD), light smokers (consuming 1-10 CPD), and heavy smokers (11 and more CPD). Never-smokers were set to have a missing value for this variable.

Lifetime QA was available for all samples. Number of quit attempts were asked only in VTSABD/TSA and MASATS. In Add Health, participants were asked directly whether or not they had tried to quit in the past six months. Smokers who answered “yes” were coded as 1, while smokers who have never tried to quit were coded as “0”. For VTSABD/TSA and MASATS, smokers who have tried to quit at least once were coded as 1, while other smokers were coded as 0. Never-smokers have a missing value for this variable. The maximum value across waves was used for each participant’s lifetime quit attempt, similar to lifetime smoking initiation.

Smoking behaviors were studied differently in VTSABD. In wave one until four, twins were only asked about lifetime smoking status and maximum cigarettes per day. In the TSA follow-up, a greater detail about smoking habit was asked. For example, questions concerning ND symptoms, patterns of smoking, and lifetime quit attempts were added in the questionnaire. Thus, information on the participants’s QA is provided only from TSA, while information on maximum CPD and SI were obtained from wave one until TSA.

Table 1 Samples and Measures Used

Sample	Age Range	Study Design	N	Years	Smoking Initiation	Smoking Amount	Quit Attempt
					Ever/Never	None: 0 CPD, Light: 1-10 CPD, Heavy: > 10 CPD	Ever/Never
ADH	11-32	Longitudinal study	1444	1994-2009	Have you ever tried cigarette smoking, even just 1 or 2 puffs?	During the past 30 days, on the days you smoked, how many cigarettes did you smoke each day?	During the past 6 months, have you tried to quit smoking cigarettes?
MASATS	11-19	Cross-Sectional	2081	1999	On how many occasions have you smoked cigarettes in your lifetime?	During the past 30 days about how many cigarettes have you smoked per day?	How many times, if any, did you try to stop smoking?
VTSABD/ TSA	8-18/ 22-32	Longitudinal study	2797	1990-2007	Have you ever smoked, even if you were trying just one cigarette?	During the past 7 days, how many cigarettes do you smoke on a typical day?	How many times in your life have you seriously tried to quit using tobacco?

Statistical Analyses

Testing age differences

To analyze the influence of age on the contribution of genetic and environmental factors to liability of QA, participants were categorized into several age groups. From the available data, three age groups could be formed: younger adolescents (age 12-15), older adolescents (age 16-

19), and young adults (age 22-32) (Table 2). Several reasons for choosing these age groups: (1) all three smoking behavior variables were collected in these age groups from at least two samples, allowing for comparison of estimates. (2) Participants aged less than 12 were collected in all samples, but the prevalence of smoking was too low. So this group was excluded. (3) Data for those aged 20-22 were also excluded because there were only collected in Add Health. (4) Creating age groups with 1-year intervals was not warranted because the resulting sample sizes, especially for MASATS, were too low for a meaningful analysis. As a comparison, another age group was created where all participants from each sample were analyzed together (All participants). Therefore, nine data groups based on age and sample were created for this study, as listed in Table 2.

Table 2 Data Groups Based on Age for This Study

Study Name	Total Number of Subjects for Each Data Group			
	All participants	Age 12-15	Age 16-19	Age 22-32
		Younger adolescents	Older adolescents	Young adults
Add Health	1444	723	1081	1304
MASATS	2081	1316	611	N/A
VTSABD/TSA	2797	N/A	N/A	1075

Descriptive statistics

To compare prevalences of smoking behaviors between males and females for all data groups, Pearson's Chi-squared Test for count data was used. The test was done with 2 degrees of freedom for CPD and 1 degrees of freedom for SI and QA. The prevalences are shown in Tables 3-5 for Add Health, MASATS, and VTSABD/TSA, respectively. The tetrachoric and polychoric correlations for the three smoking variables for each subgroup are shown in Table 6.

Table 3 Prevalence of Smoking Behaviors of by Age and Sex for Add Health

Age Group	Younger Adolescents (12-15)				Older Adolescents (16-19)				Young Adults (22-32)				All Participants				p-Value ^b				
	Males		Females		p-Value ^a	Males		Females		p-Value ^a	Males		Females		p-Value ^a						
	N	%	N	%		N	%	N	%		N	%	N	%							
Total	364	50	359	50		559	52	522	48		646	50	658	50		734	51	710	49		
Lifetime smoking					NS					*					NS					NS	NS
no	86	24	86	24		99	18	122	23		117	18	147	22		145	20	164	23		
yes	278	76	273	76		460	82	400	77		529	82	511	78		589	80	546	77		
Maximum cig/day					**					**					**					**	NS
0 (none)	75	29	101	39		116	27	142	38		145	28	188	38		146	26	189	37		
light (1-10 CPD)	117	45	117	45		173	40	162	44		206	40	214	43		228	41	230	45		
heavy (11 CPD +)	70	27	40	16		144	33	68	18		164	32	91	18		179	32	91	18		
missing	102		101			126		150			131		165			181		200			
Lifetime quit attempt					NS					NS					NS					NS	NS
no	58	32	44	28		97	30	62	26		108	29	79	26		121	30	87	27		
yes	125	68	112	72		226	70	175	74		264	71	229	74		289	70	238	73		
missing	181		203			236		285			274		350			324		385			

Note: p-Value^a: statistical testing between sexes. p-Value^b: statistical testing between age groups, NS: p-Value > 0.05, (*): p-Value < 0.05, (**): p-Value < 0.005

Table 4 Prevalence of Smoking Behaviors of by Age and Sex for MASATS

Age Group	Younger Adolescents (12-15)				Older Adolescents (16-19)				All Participants				p-Value ^b								
	Males		Females		p-Value ^a	Males		Females		p-Value ^a	Males			Females		p-Value ^a					
	N	%	N	%		N	%	N	%		N	%		N	%						
Total	541	41	775	59		230	38	381	62		846	41	1235	59							
Lifetime smoking					NS					NS					NS						**
no	411	76	572	74		123	53	184	48		605	72	834	68							
yes	130	24	203	26		107	47	197	52		241	28	401	32							
Maximum cig/day					NS					NS					NS					NS	NS
0 (none)	100	79	139	72		78	74	131	68		181	76	271	70							

light (1-10 CPD)	22	17	42	22		19	18	46	24		42	18	88	23		
heavy (11 CPD +)	5	4	13	7		9	8	15	8		14	6	28	7		
missing	414		581			124		189			609		848			
Lifetime quit attempt					NS					NS					NS	NS
no	90	72	129	68		73	70	118	63		165	71	247	66		
yes	35	28	60	32		32	30	69	37		68	29	129	34		
missing	416		586			125		194			613		859			

Note: p-Value^a: statistical testing between sexes. p-Value^b: statistical testing between age groups, NS: p-Value > 0.05, (*): p-Value < 0.05, (**): p-Value < 0.005

Table 5 Prevalence of Smoking Behaviors of by Age and Sex for VTSABD/TSA

Age Group	Younger Adults (22-32)					All Participants					p-Value ^b
	Males		Females		p-Value ^a	Males		Females		p-Value ^a	
	N	%	N	%		N	%	N	%		
Total	421	39	654	61		1286	46	1511	54		
Lifetime smoking					NS					**	**
no	81	19	149	23		440	34	657	43		
yes	340	81	505	77		846	66	854	57		
Maximum cig/day					NS					NS	**
0 (none)	188	55	298	59		500	59	547	64		
light (1-10 CPD)	82	24	119	24		222	26	195	23		
heavy (11 CPD +)	70	21	87	17		121	14	111	13		
missing	81		150			443		658			
Lifetime quit attempt					NS					NS	NS
no	161	47	247	49		161	47	247	49		
yes	179	53	258	51		179	53	258	51		
missing	81		149			946		1006			

Note: p-Value^a: statistical testing between sexes. p-Value^b: statistical testing between age groups, NS: p-Value > 0.05, (*): p-Value < 0.05, (**): p-Value < 0.005

Table 6 Tetrachoric and Polychoric Correlations for Smoking Behaviors by Sample, Age, Sex, and Zygosity

Smoking Behavior	Study Name	Age groups	Number of Twin pairs	Males				Females				DZO	(SE)
				MZ	(SE)	DZ	(SE)	MZ	(SE)	DZ	(SE)		
Smoking Initiation	Add Health	12- to 15	405	0.75	0.12	0.66	0.17	0.84	0.09	0.44	0.23	0.26	0.18
		16- to 19	566	0.76	0.10	0.51	0.19	0.58	0.13	0.62	0.14	0.32	0.16
		22- to 32	696	0.77	0.09	0.68	0.14	0.68	0.11	0.49	0.15	0.25	0.15
		All participants	732	0.75	0.09	0.52	0.16	0.68	0.10	0.53	0.14	0.40	0.12
	MASATS	12- to 15	722	0.79	0.10	0.63	0.14	1.00	Inf	1.00	Inf	0.71	0.09
		16- to 19	358	0.72	0.14	1.00	Inf	0.89	0.06	0.62	0.16	0.36	0.22
		All participants	1160	0.74	0.08	0.79	0.08	1.00	Inf	0.87	0.04	0.66	0.08
	VTSABD/TSA	22- to 32	684	0.81	0.09	0.33	0.31	0.70	0.09	0.57	0.22	0.21	0.24
All participants		1406	0.67	0.06	0.59	0.09	0.71	0.05	0.59	0.09	0.57	0.07	
Smoking Amount	Add Health	12- to 15	405	0.53	0.16	-0.22	0.20	0.78	0.09	-0.15	0.26	0.57	0.12
		16- to 19	566	0.71	0.08	0.37	0.13	0.68	0.09	0.22	0.19	0.48	0.11
		22- to 32	696	0.55	0.11	0.27	0.12	0.70	0.07	0.05	0.17	0.52	0.09
		All participants	732	0.59	0.09	0.29	0.12	0.68	0.08	0.05	0.17	0.52	0.09
	MASATS	12- to 15	722	1.00	2.82	0.56	0.38	0.54	0.22	0.39	0.27	0.45	0.27
		16- to 19	358	0.61	0.31	0.39	0.38	0.76	0.12	0.43	0.27	-1.00	2487.00
		All participants	1160	0.66	0.22	0.42	0.04	0.67	0.43	0.44	0.91	0.01	0.07
	VTSABD/TSA	22- to 32	684	0.65	0.12	-0.03	0.26	0.62	0.09	0.59	0.17	0.48	0.15
All participants		1406	0.64	0.06	0.26	0.13	0.56	0.07	0.31	0.15	0.36	0.11	
Quit Attempt	Add Health	12- to 15	405	0.99	33550.00	-0.97	5.48	-0.02	0.42	0.59	0.37	0.33	0.32
		16- to 19	566	0.26	0.25	-0.01	0.23	-0.24	0.33	0.44	0.37	0.36	0.29
		22- to 32	696	0.35	0.24	-0.02	0.22	-0.25	0.30	0.39	0.33	0.43	0.23
		All participants	732	0.30	0.23	-0.06	0.21	-0.07	0.28	0.44	0.29	0.42	0.21
	MASATS	12- to 15	722	1.E-09	0.51	-0.07	0.52	0.61	0.23	-0.33	0.32	0.70	0.23
		16- to 19	358	0.70	0.29	0.82	0.22	0.70	0.18	0.40	0.36	-0.40	0.45
		All participants	1160	0.40	0.30	0.48	0.29	0.65	0.14	-0.02	0.26	0.33	0.27
	VTSABD/TSA	22- to 32	684	0.71	0.13	0.65	0.20	0.47	0.13	0.10	0.27	0.58	0.16
All participants		1406	0.71	0.13	0.65	0.20	0.47	0.13	0.10	0.27	0.58	0.16	

Note: SE: Standard Error

Twin modeling: CCC model for SI, SA, and QA

The full model: CCC-Cholesky ACE model with qualitative and quantitative sex differences (Model 1)

The full model consists of all parameters tested in each data groups: the ACE influences on each smoking phenotype, causal pathways, and sex differences (Figure 5). A CCC model was chosen because this study used a binary SI phenotype and observations for SA and QA can only be obtained from smokers. Five zygosity groups based on sex were used in this study: monozygotic males (MZM), monozygotic females (MZF), dizygotic males (DZM), dizygotic females (DZF), and opposite sex dizygotic twins (DZO). The inclusion of DZOs allows for testing qualitative and quantitative sex differences. Only genetic correlations were estimated in this study.

A total of 11 models were tested to investigate the importance of each pathway. The models are listed in Table 7. Each model has different pathway combinations: (1) Models 2 until 4 tested the presence of sex differences in the relationship between SI, SA, and QA. (2) Models 5 until 7 tested the importance of the causal pathways. (3) Models 8 until 11 tested the importance of A and C. The models were compared and model testing was done sequentially. One model's testing results determined which sub-model to be tested next. As a result, some models were tested in one data group, but not in others (Table 7).

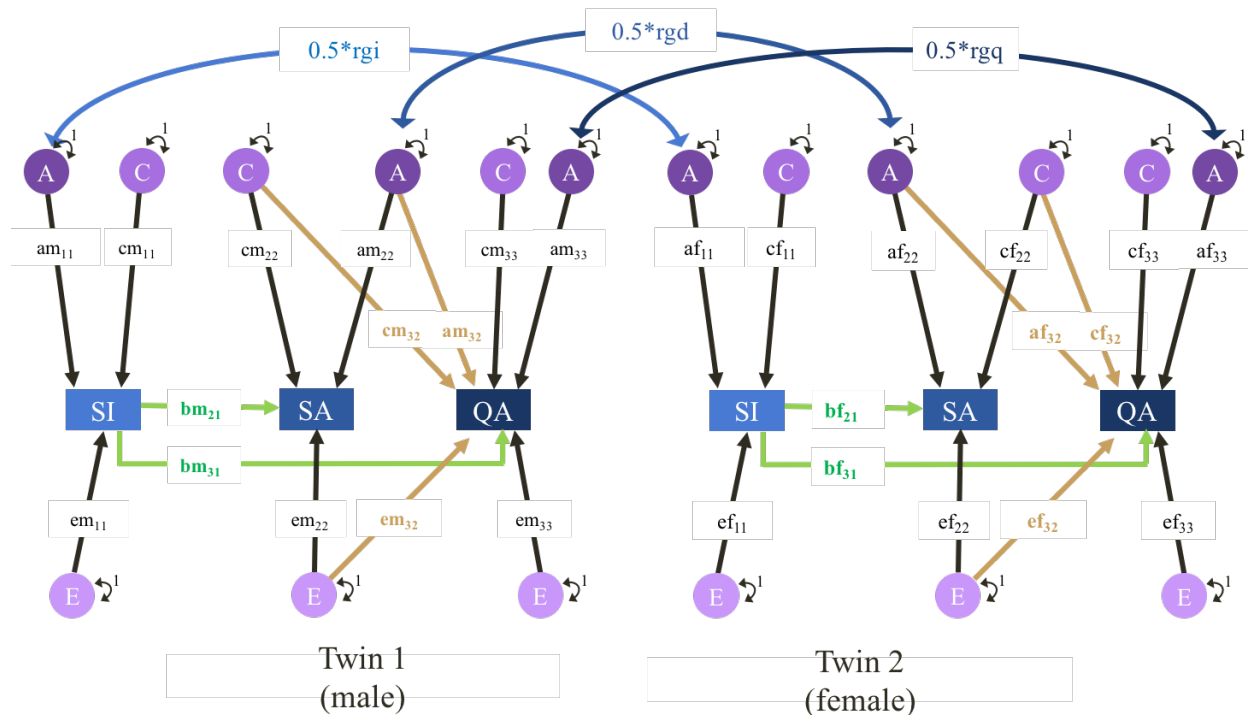


Figure 5. Model 1: CCC ACE Model With Qualitative and Quantitative Sex Differences. This model is pictured for the pathways between DZ opposite sex pair twins. The three observed smoking phenotypes are illustrated in blue boxes: SI for smoking initiation (lifetime cigarette use), SA for smoking amount (cigarettes per day, CPD), and QA for lifetime quit attempt. The genetic and environmental factors are pictured in different shades of purple circles: A for additive genetic factors, C for shared environmental factors, and E for unique environmental factors. In this model, the factors specific for each stage were estimated and influenced the corresponding smoking stage through black pathways. For example, the am_{33} path illustrates the genetic factors specifically for QA in males. The green pathways are causal pathways. The brown pathways are for the covariance pathways between SA and QA. The blue double arrow curves between each A of males and females illustrate the genetic correlation for each smoking phenotype (qualitative sex differences). Quantitative sex differences were tested by allowing different estimations for males and females for all pathways

Table 7 The 11 Models Testing The Significance of Pathways Estimates

Purpose	Model Description	Data groups									
		Add Health				MASATS			VTSABD		
		12-15	16-19	22-32	all	12-15	16-19	all	22-32	all	
Full model	1. ACE with sex differences	b	b	b	b	b	b	b	b	b	
Testing qualitative sex differences	2. All $rg = 1$	b	b	b	b	b	b	b	b	b	

Testing quantitative sex differences	3. Equate thresholds	a	a	a	a	a	a	a	a	a
	4. All $rg = 1$, equate ACE and b estimates	b	b	b	b	b	b	b	b	*
Testing causal pathways	5. Drop all causal pathways	a	a	a	a	a	a	a	a	a
	6. Drop causal pathway from SI to SA (b_{21})	a	a	a	a	a	a	a	a	a
	7. Drop causal pathway from SI to QA (b_{31})	b	b	b	b	b	b	*	*	a
Testing genetic and environmental factor pathways	8. Drop b_{31} and all c pathways	*	*	*	*	a	*	a	a	
	9. Drop b_{31} and all a pathways	a	a	a	a	*	a	a	a	
	10. Drop all c pathways									a
	11. Drop all a pathways									a

Note: green (*) cells: best fitting models; red (a) cells: the model was tested in this data group and was significantly worse than its base model; blue (b) cells: the model was tested in this data group and was not significantly worse than its base model; grey cells: the model was not tested in this data group.

Model 2: Testing qualitative sex differences

This models had the same parameters as model 1, except for the genetic correlations. The model sets genetic correlations (rg) into 1 to test the nature of genetic factors influencing smoking behaviors. If this model is not significantly different from the full model, then males and females have the same set of smoking-related genes. In other words, no qualitative sex differences are detected for smoking-related genetic factors.

Models 3-4: Testing quantitative sex differences

These models also have the same parameters as model 1, except for their variable estimates across sex. In these models, the estimates were equated across males and females. If these models were not significantly different from the full model, then there were no quantitative sex differences for the path estimates.

In model 3, threshold estimates were set equal across sexes. In the full model, different threshold estimates are allowed between males and females to model sex differences in prevalences. If model 3 is not significant, then males and females have the same threshold estimate for smoking phenotypes. In model 4, the all rgfs were set to be one and all path estimates were equated across males and females.

Models 5-7: Testing causal pathways

Models 5 to 7 were used to test the significance of causal pathways from SI to SA (b_{21}) and SI to QA (b_{31}). The base model for these models was the model without sex differences (model 4). If these models were not significantly different from the base model, then smoking progression variables were independent from SI. In model 5, b_{21} and b_{31} were set to zero, i.e. dropped from the model. In model 6, only b_{21} was dropped from the model. In model 7, only b_{31} was dropped from the model.

Models 8-11: Testing A and C

In these models, genetic and environmental pathways were removed by setting the estimate to zero. The purpose was to examine whether the factors were significantly important in explaining the liabilities for smoking behaviors. The base model for these submodels was the model without sex differences (model 4). In models 8 and 10, the shared environmental pathways were dropped from the models. In models 9 and 11, the genetic pathways were

dropped from the models. In models 8 and 9, in addition to the genetic and environmental factors, the causal pathway b_{31} was dropped from the model. In models 10 and 11, all causal pathways were present.

Model testing

The first model testing used was a likelihood ratio test, because these 11 models are nested (Neyman and Pearson 1928; Vuong 1989). The likelihood ratio test provides information on whether a submodel is significantly different than its base model or not. The test statistic is the difference in the $-2 \log$ -likelihood values between the models. Then, a p-value was calculated using a Chi-square distribution with degrees of freedom (df) equal to the difference of free parameters between the base and nested submodels. If the resulting p-value is less than 0.05, then the submodel is significantly different than its base model. This means that a certain manipulation of pathways created a worse data fit than the base model. This likelihood ratio test is useful in deciding which pathways are essential for the model to properly explain the data. Non-significant sub-models are then candidates for becoming the best-fitting model.

To choose between non-significant models, the Akaike Information Criterion (AIC) values was used (Akaike 1987; Williams and Holahan 1994). AIC is founded on information theory and helps to find the model with minimum information loss. This measure rewards goodness of fit, but also includes a higher penalty with increasing number of estimated parameters to discourage overfitting. So this method is an index of model fit and parsimony. The preferred model has the lowest AIC value. Therefore, the best fitting model is the non-significant submodel with the lowest AIC value. The estimates from best fitting models were determined and 95% confidence intervals were calculated.

Calculating proportion of variance components

Estimating the proportion of variances was done only for best fitting models. The equations for calculating for the total variance for each smoking behavior is shown in Table 8.

Software

All data analysis was done using the OpenMx package (Neale et al. 2016; Pritikin, Hunter, and Boker 2015; Boker et al. 2016) version 2.6.9 with NPSOL optimizer, in R version 3.3.1. (R Core Team 2016). Packages used for analysis are ‘psych’ (Revelle 2016) and ‘polycor’ (Fox 2016) packages.

Table 8 Equations for Calculating Proportion of Variances

Equations	
SI	
Total A variance	$A_{SI} = a_{11}^2$
Total C variance	$C_{SI} = c_{11}^2$
Total E variance	$E_{SI} = e_{11}^2$
SA	
Total A variance	$A_{SA} = a_{11}^2 \times b_{21}^2 + a_{22}^2$
A explained by SI	$A_{SA_shared} = (a_{11}^2 \times b_{21}^2) \div A_{SA}$
A unique to SA	$A_{SA_unique} = 100\% - A_{SA_shared}$
Total C variance	$C_{SA} = c_{11}^2 \times b_{21}^2 + c_{22}^2$
C explained by SI	$C_{SA_shared} = (c_{11}^2 \times b_{21}^2) \div C_{SA}$
C unique to SA	$C_{SA_unique} = 100\% - C_{SA_shared}$
Total E variance	$E_{SA} = e_{11}^2 \times b_{21}^2 + e_{22}^2$
E explained by SI	$E_{SA_shared} = (e_{11}^2 \times b_{21}^2) \div E_{SA}$
E unique to SA	$E_{SA_unique} = 100\% - E_{SA_shared}$
QA	
Total A variance	$A_{QA} = a_{11}^2 \times b_{31}^2 + a_{32}^2 + a_{33}^2$
A explained by SI	$A_{QA_sharedSI} = (a_{11}^2 \times b_{31}^2) \div A_{QA}$
A explained by SA	$A_{QA_sharedSA} = (a_{32}^2) \div A_{QA}$
A unique for QA	$A_{QA_unique} = 100\% - A_{QA_shared}$
Total C variance	$C_{QA} = c_{11}^2 \times b_{31}^2 + c_{32}^2 + c_{33}^2$
C explained by SI	$C_{QA_sharedSI} = (c_{11}^2 \times b_{31}^2) \div C_{QA}$

C explained by SA	$C_{QA_sharedSA} = (c_{32}^2) \div C_{QA}$
C unique for QA	$C_{QA_unique} = 100\% - C_{QA_shared}$
Total E variance	$E_{QA} = e_{11}^2 \times b_{31}^2 + e_{32}^2 + e_{33}^2$
E explained by SI	$E_{QA_shared} = (e_{11}^2 \times b_{31}^2) \div E_{QA}$
E explained by SA	$E_{QA_sharedSA} = (e_{32}^2) \div E_{QA}$
E unique for QA	$E_{QA_unique} = 100\% - E_{QA_shared}$

Results

Summary Statistics

There were no sex and age differences for the prevalence of QA, as shown in Tables 3 to 5. However, there were significant sample differences for QA prevalence ($p < 0.0001$). For other smoking behaviors, sex, age, and sample differences were found. Sex differences were seen for SA in all age groups of Add Health and for SI in VTSABD/TSA participants when analyzed together ($p < 0.0001$). Age differences were seen for SI between the age groups of MASATS ($p < 0.0001$) and for SI and SA between age groups of VTSABD/TSA (SI: $p < 0.0001$; SA: $p < 0.01$). Highly significant sample differences were also found for SI and SA (SI: $p = 0$; SA: $p < 0.0001$).

Table 6 lists the MZ and DZ correlations for the smoking behaviors in this study. Overall, MZ pairs have higher correlations than DZ pairs for SI and SA, across all ages and sex. However, the difference between MZ and DZ correlations were inconsistent across data groups. Some data groups have MZ correlations that were twice the DZ correlations, in others DZ correlations were greater than half of MZ correlations. This indicates possible A and C contributions for SI and SA. The correlations for QA were more difficult to interpret, where MZ pairs have higher correlations than DZ pairs in some data groups, but lower in others. Negative correlations were found for DZ males and MZ females for QA. However, the standard error of these estimates were large (Table 6).

Genetic Analysis

Table 7 lists all the models tested in this study. There were sample differences where some models fitted the data well in some data groups, but not in others. For example, model 7 was significantly different than its base model only when all VTSABD/TSA participants were analyzed together. So, this data group needed other models to test genetic and environmental factor pathways, which were models 10 and 11. These additional models were not tested in the other eight data groups and were indicated as gray in Table 7. The red color or ‘a’ label indicates that the model was tested in the particular data group, but was significantly worse than its base model. If it was not significantly worse, the model is labeled as blue or ‘b’. For example, the model 7 was significantly worse in the VTSABD/TSA all participants group, thus labeled as red, but not significantly worse in other groups, thus labeled as blue in other cells. Best fitting models were indicated as green or ‘*’ in Table 7.

Testing sex differences

Overall, there were no sex differences in the contributions of genetic and environmental on liability for smoking behaviors. There were sex differences for threshold estimates, indicating significant sex differences for the prevalence of these smoking behaviors. In all data groups, there were no qualitative sex differences for smoking behaviors. The genetic correlations between males and females for the three smoking phenotypes could be equated to one without significant loss of fit (model 2). Additionally, there were no quantitative sex differences detected, except for the thresholds. The path estimates could be equated for males and females without significant loss of model fit (model 4).

Testing causal pathways

Consistent results were also seen in models testing common pathways. Dropping all causal pathways resulted in significant loss of fit (model 5). Dropping only the causal pathway from SI to SA also worsened the model fit (model 6), but not when the causal pathway from SI to QA was dropped (model 7). This suggests the importance of shared liabilities between SI and SA and separate liabilities between SI and QA. However, when VTSABD/TSA all participants were analyzed together, all causal pathways needed to be kept in the model.

Testing A and C and determining best fitting models

Add Health

Across all Add Health age groups (Table 9), C could be dropped from the model without significant loss of fit. Dropping A made the model significantly worse than the base model. Based on model testing results and AIC values, the best fitting model for all Add Health groups was the AE model with the causal pathway from SI to SA (b_{21}).

Table 9 Model Comparison by Age Group: Add Health Subjects

Age group	Model Description		ep	-2LL	df	AIC	Δ LL	Δ df	p
	Base	Comparison							
Younger adolescents (12-15)	Full model	Full model	39	2179.14	1549	-918.86	-	-	-
		All rg = 1	36	2179.18	1552	-924.82	0.04	3	0.998
		Equate thresholds	34	2200.15	1554	-907.85	21.01	5	8.E-04
		All rg = 1, equate ACE and b estimates	22	2188.17	1566	-943.83	9.03	17	0.939
	ACE with no sex differences	Drop all causal pathways	20	2200.46	1568	-935.54	12.28	2	0.002
		Drop causal pathway from SI to SA (b_{21})	21	2199.26	1567	-934.74	11.09	1	9.E-04
		Drop causal pathway from SI to QA (b_{31})	21	2188.39	1567	-945.61	0.22	1	0.64
		Drop b_{31} and all c pathways	17	2188.65	1571	-953.35	0.48	5	0.993
	Drop b_{31} and all a pathways	17	2203.72	1571	-938.28	15.55	5	0.008	
Older adolescents (12-15)	Full model	Full model	39	3294.85	2413	-1531.15	-	-	-
		All rg = 1	36	3295.79	2416	-1536.21	0.94	3	0.82
		Equate thresholds	34	3330.39	2418	-1505.61	35.54	5	1.17E-06
		All rg = 1, equate ACE and b estimates	22	3300.53	2430	-1559.47	5.69	17	0.995

	ACE with no sex differences	Drop all causal pathways	20	3325.68	2432	-1538.32	25.15	2	3.46E-06
		Drop causal pathway from SI to SA (b_{21})	21	3324.82	2431	-1537.18	24.28	1	8.32E-07
		Drop causal pathway from SI to QA (b_{31})	21	3300.78	2431	-1561.22	0.25	1	0.62
		Drop b_{31} and all c pathways	17	3305.45	2435	-1564.55	4.92	5	0.43
		Drop b_{31} and all a pathways	17	3315.02	2435	-1554.98	14.49	5	0.01
Young adults (22-32)	Full model	Full model	39	4051.23	2959	-1866.77	-	-	-
		All $rg = 1$	36	4051.23	2962	-1872.77	-0.0001	3	1
		Equate thresholds	34	4090.10	2964	-1837.90	38.87	5	2.52E-07
		All $rg = 1$, equate ACE and b estimates	22	4065.26	2976	-1886.74	14.03	17	0.66
	ACE with no sex differences	Drop all causal pathways	20	4094.76	2978	-1861.24	29.49	2	3.94E-07
		Drop causal pathway from SI to SA (b_{21})	21	4093.34	2977	-1860.66	28.07	1	1.17E-07
		Drop causal pathway from SI to QA (b_{31})	21	4065.73	2977	-1888.27	0.46	1	0.50
		Drop b_{31} and all c pathways	17	4068.21	2981	-1893.79	2.94	5	0.71
		Drop b_{31} and all a pathways	17	4083.64	2981	-1878.36	18.38	5	0.003
All participants	Full model	Full model	39	4386.31	3209	-2031.69	-	-	-
		All $rg = 1$	36	4386.31	3212	-2037.69	0	3	1
		Equate thresholds	34	4438.22	3214	-1989.78	51.91	5	5.63E-10
		All $rg = 1$, equate ACE and b estimates	22	4396.32	3226	-2055.68	10.01	17	0.90
	ACE with no sex differences	Drop all causal pathways	20	4430.61	3228	-2025.39	34.30	2	3.57E-08
		Drop causal pathway from SI to SA (b_{21})	21	4429.93	3227	-2024.07	33.61	1	6.74E-09
		Drop causal pathway from SI to QA (b_{31})	21	4397.66	3227	-2056.34	1.34	1	0.25
		Drop b_{31} and all c pathways	17	4401.88	3231	-2060.12	5.56	5	0.35
		Drop b_{31} and all a pathways	17	4415.87	3231	-2046.13	19.55	5	0.002

Note: ep: number of estimated parameters; AIC: Akaike Information Criterion. Best fitting models are bolded.

MASATS

Different results were found between the age groups from MASATS when A or C were dropped from the sex model (Table 10). Dropping C greatly reduced model fit in younger adolescents, but A could be dropped without significant loss in model fit. In contrast, dropping C in the model for older adolescents did not cause any significant fit loss, while dropping A did. However, when all MASATS participants were analyzed together, A and C were equally important. So, the best fitting model for younger adolescents was the CE model, for older

adolescents was the AE model and for MASATS all participants was the ACE model. All best fitting models for MASATS have a causal pathway from SI to SA (b_{21}).

Table 10 Model Comparison by Age Group: MASATS Subjects

Age group	Model Description		ep	-2LL	df	AIC	Δ LL	Δ df	p
	Base	Comparison							
Younger adolescents (12-15)	Full model	Full model	39	2043.13	1939	-1834.87	-	-	-
		All rg = 1	36	2043.00	1942	-1841.00	-0.13	3	1
		Equate thresholds	34	2057.22	1944	-1830.78	14.09	5	0.02
		All rg = 1, equate ACE and b estimates	22	2062.23	1956	-1849.77	19.11	17	0.32
		Drop b₃₁ and all a pathways	17	2069.76	1961	-1852.24	7.52	5	0.18
	ACE with no sex differences	Drop all causal pathways	20	2068.83	1958	-1847.17	6.59	2	0.04
		Drop causal pathway from SI to SA (b_{21})	21	2068.70	1957	-1845.30	6.46	1	0.01
		Drop causal pathway from SI to QA (b_{31})	21	2062.21	1957	-1851.79	-0.02	1	1
		Drop b_{31} and all c pathways	17	2090.81	1961	-1831.19	28.58	5	2.81E-05
		Drop b₃₁ and all a pathways	17	2069.76	1961	-1852.24	7.52	5	0.18
Older adolescents (12-15)	Full model	Full model	39	1495.86	1168.00	-840.14	-	-	-
		All rg = 1	36	1495.29	1171.00	-846.71	-0.57	3	1.00
		Equate thresholds	34	1515.65	1173.00	-830.35	19.79	5	0.001
		All rg = 1, equate ACE and b estimates	22	1513.30	1185.00	-856.70	17.44	17	0.42
		Drop b₃₁ and all a pathways	17	1516.49	1190.00	-863.51	3.18	5	0.67
	ACE with no sex differences	Drop all causal pathways	20	1520.45	1187.00	-853.55	7.15	2	0.03
		Drop causal pathway from SI to SA (b_{21})	21	1520.08	1186.00	-851.92	6.78	1	0.01
		Drop causal pathway from SI to QA (b_{31})	21	1514.36	1186.00	-857.64	1.06	1	0.30
		Drop b₃₁ and all c pathways	17	1516.49	1190.00	-863.51	3.18	5	0.67
		Drop b_{31} and all a pathways	17	1525.29	1190.00	-854.71	11.99	5	0.03
All Subjects	Full model	Full model	39	3771.69	3317.00	-2862.31	-	-	-
		All rg = 1	36	3774.39	3320.00	-2865.61	2.70	3	0.44
		Equate thresholds	34	3800.26	3322.00	-2843.74	28.57	5	2.81E-05
		All rg = 1, equate ACE and b estimates	22	3796.19	3334.00	-2871.81	24.50	17	0.11
		Drop b₃₁ and all a pathways	17	3796.68	3335.00	-2873.32	0.49	1	0.48
	ACE with no sex differences	Drop all causal pathways	20	3809.79	3336.00	-2862.21	13.61	2	0.001
		Drop causal pathway from SI to SA (b_{21})	21	3809.71	3335.00	-2860.29	13.53	1	2.E-04
		Drop causal pathway from SI to QA (b_{31})	21	3796.68	3335.00	-2873.32	0.49	1	0.48
		Drop b_{31} and all c pathways	17	3835.17	3339.00	-2842.83	38.99	5	2.39E-07
		Drop b_{31} and all a pathways	17	3812.54	3339.00	-2865.46	16.36	5	0.006

Note: ep: number of estimated parameters; AIC: Akaike Information Criterion. Best fitting models are bolded.

VTSABD/TSA

The fit statistics for models tested in VTSABD/TSA data groups are shown in Table 11.

In both data groups, the A and C pathways were important because dropping them resulted in

significant loss of fit. The best fitting model for VTSABD/TSA young adults was the ACE model with only the b_{21} causal pathway. When VTSABD/TSA participants were analyzed together, the best fitting model was an ACE model with all causal pathways (b_{21} and b_{31}).

Table 11 Model Comparison by Age Group: VTSABD/TSA Subjects

Age group	Model Description		ep	-2LL	df	AIC	Δ LL	Δ df	p
	Base	Comparison							
Young adults (22-32)	Full model	Full model	39	3633.89	2773.00	-1912.11	-	-	-
		All rg = 1	36	3634.98	2776.00	-1917.02	1.09	3	0.78
		Equate thresholds	34	3665.00	2778.00	-1891.00	31.11	5	8.90E-06
		All rg = 1, equate ACE and b estimates	22	3649.68	2790.00	-1930.32	15.80	17	0.54
	ACE with no sex differences	Drop all causal pathways	20	3665.23	2792.00	-1918.77	15.54	2	4.E-04
		Drop causal pathway from SI to SA (b_{21})	21	3665.19	2791.00	-1916.81	15.50	1	8.25E-05
		Drop causal pathway from SI to QA (b_{31})	21	3649.81	2791.00	-1932.19	0.13	1	0.72
		Drop b_{31} and all c pathways	17	3664.11	2795.00	-1925.89	14.43	5	0.01
		Drop b_{31} and all a pathways	17	3669.21	2795.00	-1920.79	19.53	5	0.002
		All Subjects	Full model	39	7399.88	5347.00	-3294.12	NA	NA
All Subjects	Full model	All rg = 1	36	7399.83	5350.00	-3300.17	-0.05	3	1
		Equate thresholds	34	7516.65	5352.00	-3187.35	116.77	5	1.51E-23
		All rg = 1, equate ACE and b estimates	22	7416.51	5364.00	-3311.49	16.63	17	0.48
		Drop all causal pathways	20	7474.25	5366.00	-3257.75	57.75	2	2.89E-13
	ACE with no sex differences	Drop causal pathway from SI to SA (b_{21})	21	7470.41	5365.00	-3259.59	53.90	1	2.11E-13
		Drop causal pathway from SI to QA (b_{31})	21	7430.37	5365.00	-3299.63	13.86	1	2.E-04
		Drop all c pathways	17	7431.05	5368.00	-3304.95	14.54	4	6.E-03
		Drop all a pathways	17	7437.53	5368.00	-3298.47	21.02	4	3.E-04

Note: ep: number of estimated parameters; AIC: Akaike Information Criterion. Best fitting models are bolded.

Path estimates for ACE models

The path estimates from the ACE models without quantitative sex differences for each data group are presented in Table 12. The significant paths based on model testing are bolded in the table. There were a number of paths with high estimates, but non-significant. For example, the additive genetic paths in MASATS younger adolescents (Table 12). The a paths for SI, SA and the covariance a path between SA and QA had high estimates (0.4-0.63), but they were not

significant, based on model testing. Similarly for the causal pathway from SI to QA (b_{31}). It had a high estimate of 0.41 in MASATS older adolescents, but was not significant. In other age groups, the b_{31} paths had low estimates (b_{31} : -0.11-0.16) and were non-significant. This path was only significant in VTSABD/TSA analyzed together (b_{31} : 0.56) (Table 12).

Path estimates for best fitting models

The path estimates from best fitting models are presented in Figures 6-14. The estimates in Figures 6-14 differ from Table 12 because best fitting models have less paths than the ACE models. In all data groups, the causal pathways from SI to SA were large and significant (b_{21} =0.72-0.89). The causal pathway from SI to QA was only significant and large ($b_{31} = 0.63$) in VTSABD/TSA participants when analyzed together. Significant and large path estimates for covariance pathways ($a_{32} = 0.08$ -0.63, $c_{32} = 0.26$ -0.69, $e_{32} = 0.07$ -0.71) suggest shared liabilities between SA and QA (Figures 6-14). The path estimates from best fitting models were used to calculate the variance component estimates shown in Table 13.

Variance component estimates for QA from best fitting models

Add Health

Genetic factors accounted for a small proportion of the total variance for lifetime quit attempt ($A = 20$ -24%), after accounting for SI and SA (Table 13). There was a slight increase of A influence across age groups in Add Health. In addition, there were genetic factors unique to QA in younger adolescents ($A = 15\%$), but in older age groups, the A for QA was shared completely with SA. The remaining variance of QA was accounted for by E ($E = 76$ -80%).

MASATS

A and C were significant for QA in MASATS, but had differing importance across age groups. C was important for QA (C= 28%) in younger adolescents, but not in older adolescents (C = 0%). Instead, A was significant (A = 61%) for QA in older adolescents (Table 13). When the participants were analyzed together, A and C were both important for QA (A = 47%, C = 7%). The remaining variance was accounted for E (E = 39-72%). Most of the genetic and environmental factors for QA were shared with SA (34-100%), but not with SI.

VTSABD/TSA

In VTSABD/TSA, genetic and environmental factors had significant influences for QA. In young adults, C and E accounted for a larger proportion of the total variance for QA than A (A = 3%, C = 48%, E= 49%). The factors for QA were shared with SA, but not with SI. However, when the participants were analyzed together, genetic factors accounted for a higher proportion of the variance for QA, while the shared and unique environmental factors were still important for QA (A = 26%, C = 33%, E = 41%). Additionally, the genetic and environmental factors for QA were shared partially with SI (27-54%).

Table 12 Path Coefficients and 95% Confidence Intervals of ACE Models Without Quantitative Sex Differences for Each Data Group

Dataset		Add Health				MASATS			VTSABD/TSA	
		12-15 (95% CI)	16-19 (95% CI)	22-32 (95% CI)	All (95% CI)	12-15 (95% CI)	16-19 (95% CI)	All (95% CI)	22-32 (95% CI)	All (95% CI)
Smoking Initiation	a ₁₁	0.83	0.81	0.81	0.78	0.51	-0.77	0.52	0.83	0.60
		NA	(0.63-0.88)	(0.57-0.88)	NA	(0.29-0.62)	NA	(0.31-0.56)	NA	(0.55-0.64)
	c ₁₁	0.32	0.24	0.26	0.35	0.82	0.52	0.79	-2.00E-04	0.60
	e ₁₁	NA	(-0.55-0.6)	(-0.55-0.51)	(0.34-0.36)	(0.73-0.91)	(0.42-0.56)	(0.78-0.84)	NA	(0.45-0.64)
		0.46	0.54	0.53	0.54	0.28	0.38	0.31	0.56	0.53
		(0.34-0.55)	(0.44-0.64)	(0.44-0.63)	NA	(-0.27-0.32)	(0.3-0.39)	(0.29-0.32)	NA	(0.5-0.57)
Causal Pathways	b ₂₁	0.74	0.88	0.87	0.88	0.76	0.78	0.74	0.81	0.81
		(0.64-0.8)	(0.82-0.92)	(0.62-0.94)	NA	(0.71-0.77)	(0.74-0.79)	(0.71-0.76)	NA	(0.8-0.83)
	b ₃₁	-0.11	0.09	0.11	0.16	0.07	0.41	0.16	0.16	0.63
		(-0.49-0.21)	(-0.14-0.42)	(-0.06-0.28)	NA	(0.05-0.17)	(0.38-0.41)	(0.11-0.28)	NA	(0.56-0.66)
Smoking Amount	a ₂₂	0.53	0.09	0.25	0.21	0.4	0.45	0.48	-0.01	0.42
		(0.15-0.56)	(-0.38-0.52)	(-0.4-0.57)	(0.09-0.33)	(0.03-0.73)	(0.42-0.48)	(0.37-0.52)	NA	(0.39-0.51)
	c ₂₂	-1.00E-03	0.4	0.29	0.29	0.25	0.03	0.02	0.5	0.21
		(-0.07-0.37)	(0.17-0.54)	(-0.38-0.46)	(0.26-0.4)	(-0.06-0.33)	(-0.14-0.06)	(-0.04-0.09)	NA	(0.15-0.22)
	e ₂₂	0.42	0.24	0.3	0.25	0.45	0.44	0.46	0.31	0.34
		(0.16-0.46)	(0.16-0.48)	(-0.07-0.5)	(0.21-0.29)	NA	(0.4-0.46)	(0.4-0.49)	NA	(0.3-0.36)
Covariance	a ₃₂	0.45	0.02	0.24	0.18	0.63	0.52	0.61	0.21	0.17
		(0.42-0.47)	NA	(0-0.45)	(0-0.19)	NA	NA	(0.57-0.64)	NA	(0.11-0.23)
	c ₃₂	1.00E-04	0.45	0.39	0.41	0.16	0.52	0.17	0.65	0.44
		(0-0.29)	(0.08-0.63)	NA	NA	(-0.02-0.17)	NA	(0.04-0.5)	NA	(0.4-0.51)
	e ₃₂	0.65	0.43	0.41	0.51	0.42	0.26	0.37	0.48	0.07
		NA	(0-0.87)	(0-0.96)	NA	(0.33-0.51)	(0.24-0.27)	NA	NA	(0-0.27)
Quit Attempts	a ₃₃	1.00E-04	1.00E-04	1.00E-04	0.02	0.07	0.15	0.37	0.19	0.30
		NA	NA	NA	NA	NA	NA	(0.27-0.58)	NA	NA
	c ₃₃	1.00E-04	1.00E-04	1.00E-04	1.00E-04	1.00E-04	0.02	0.09	0.09	0.00
		NA	NA	NA	NA	NA	NA	(0-0.28)	NA	NA
	e ₃₃	0.6	0.78	0.78	0.72	0.62	0.46	0.55	0.48	0.54
		NA	(0.6-0.94)	(0.64-0.92)	NA	(0.61-0.75)	NA	(0.51-0.58)	NA	(0.47-0.57)

Note: Significant paths are bolded.

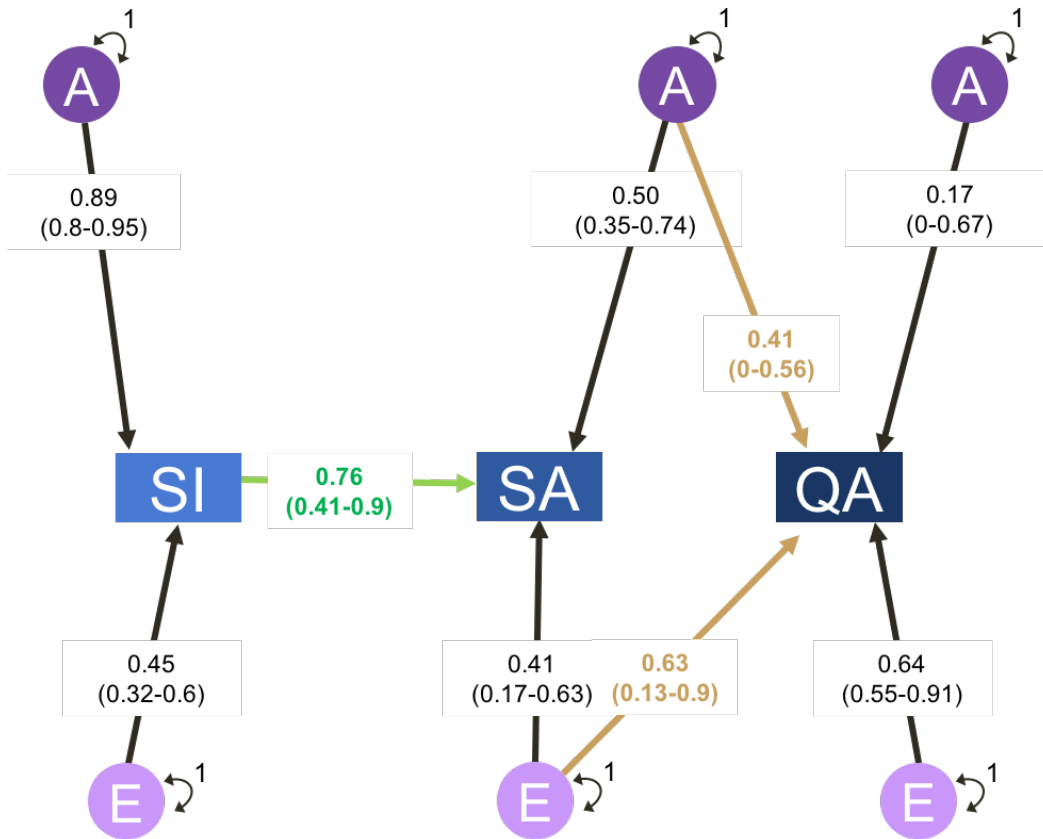


Figure 6. The Best Fitting Model and Path Estimates for Younger Adolescents (age 12-15): Add Health. Note: 95% Confidence Intervals are in brackets.

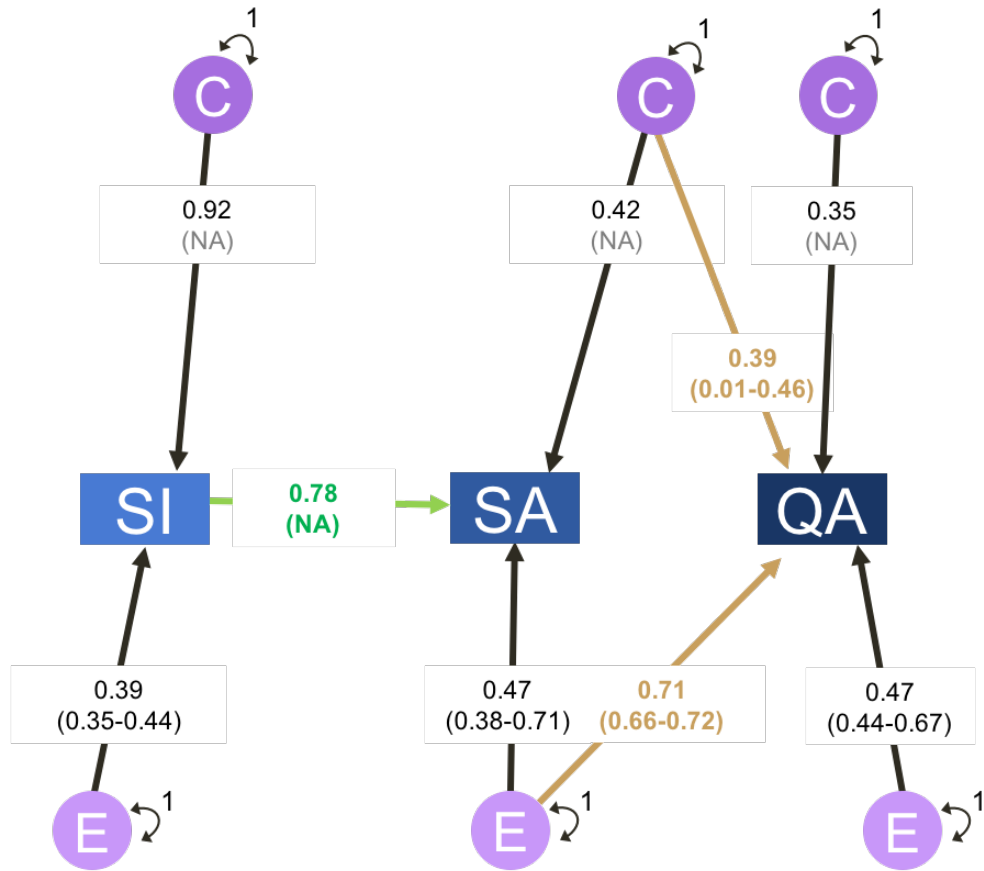


Figure 7. The Best Fitting Model and Path Estimates for Younger Adolescents (age 12-15): MASATS. Note: 95% Confidence Intervals are in brackets.

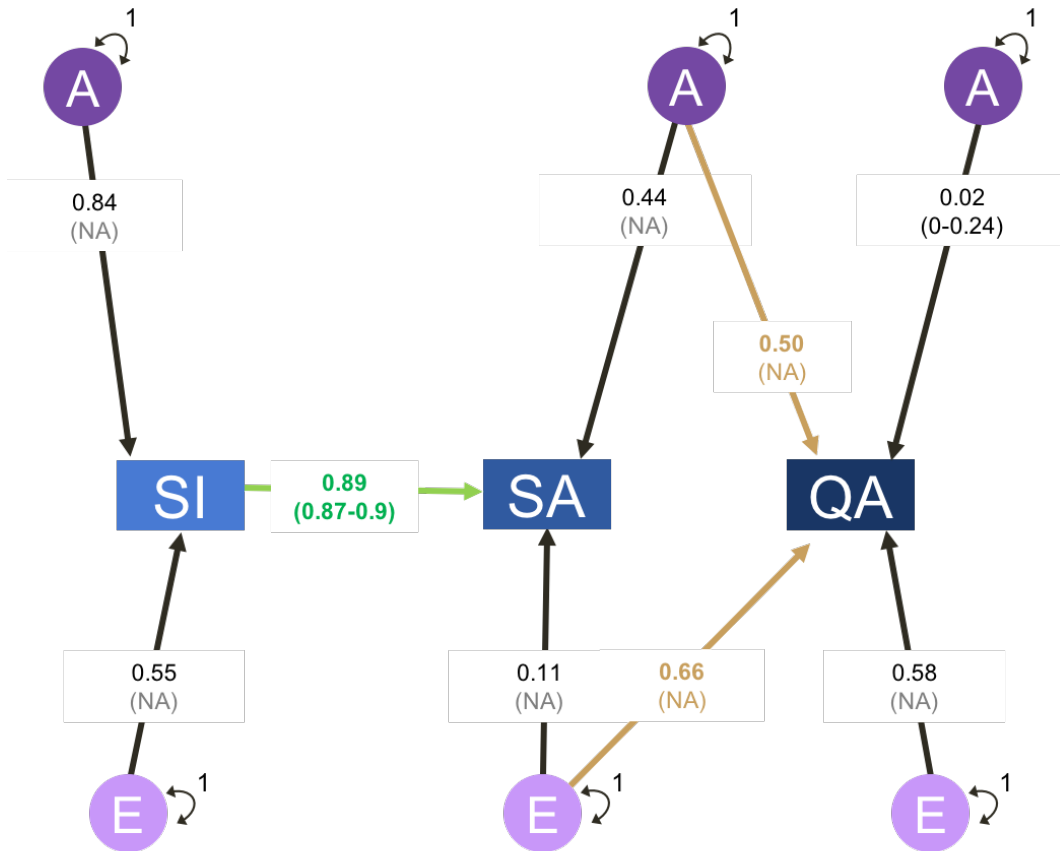


Figure 8. The Best Fitting Model and Path Estimates for Older Adolescents (age 16-19): Add Health. Note: 95% Confidence Intervals are in brackets.

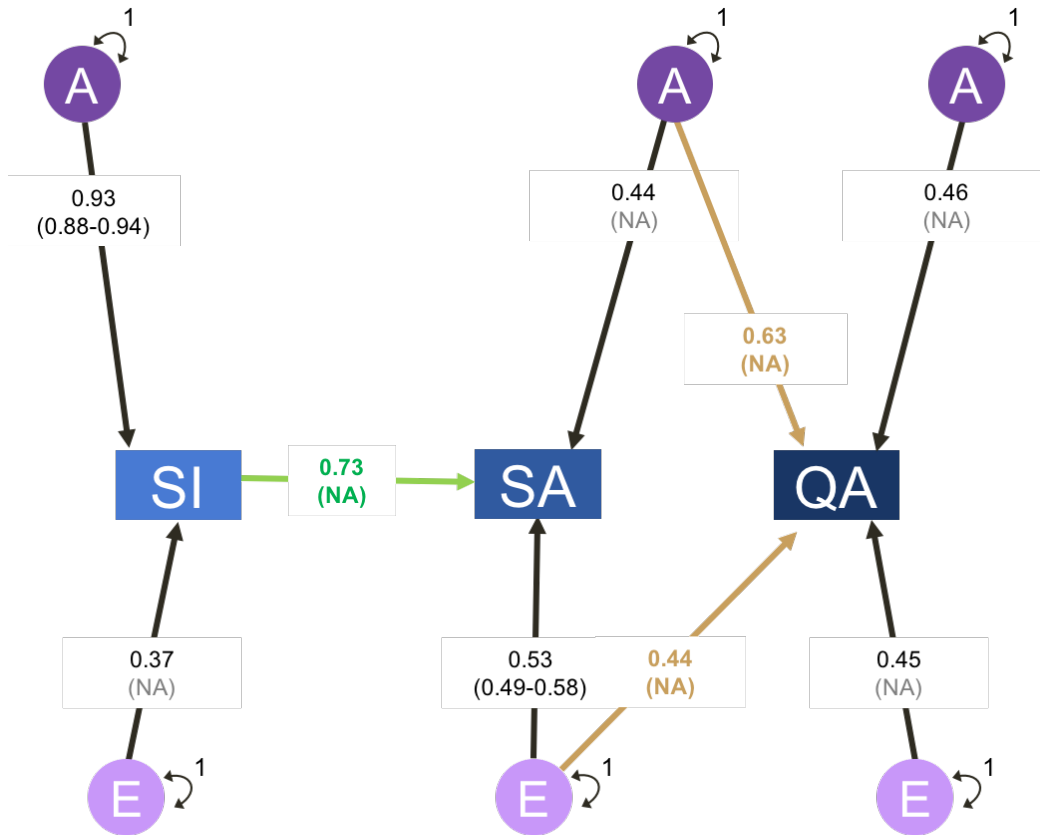


Figure 9. The Best Fitting Model and Path Estimates for Older Adolescents (age 16-19): MASATS. Note: 95% Confidence Intervals are in brackets.

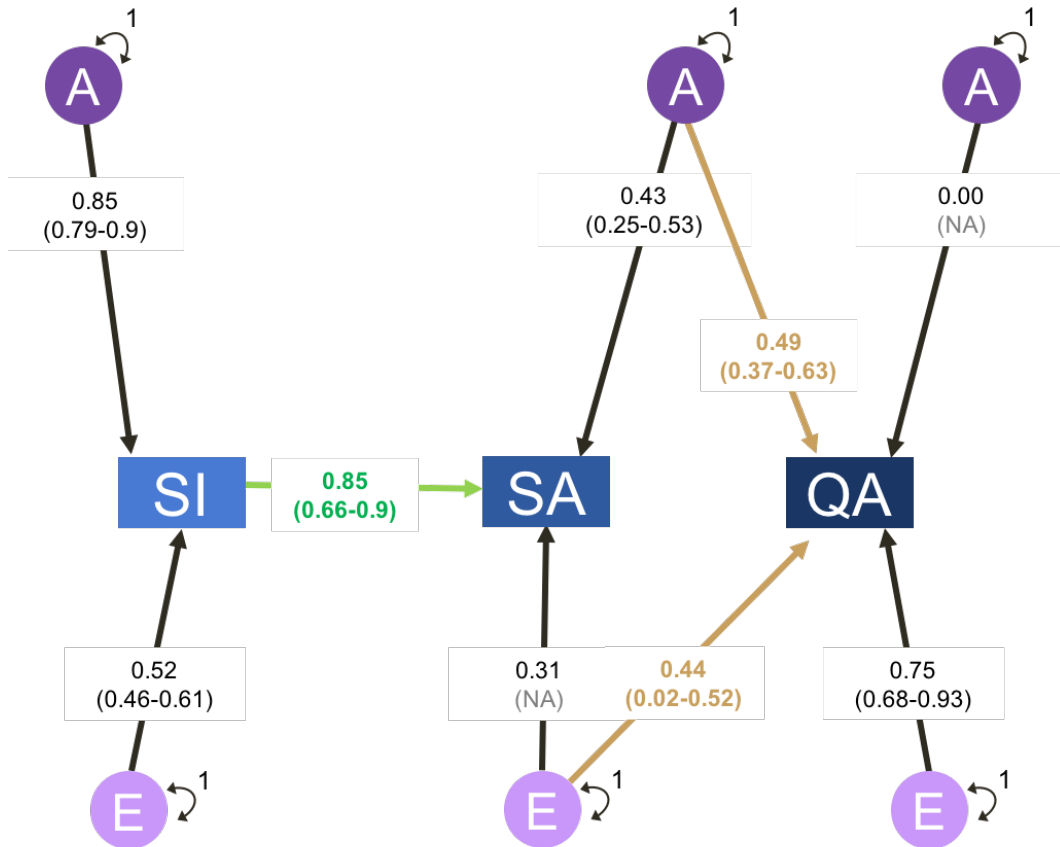


Figure 10. The Best Fitting Model and Path Estimates for Young Adults (age 22-32): Add Health. Note: 95% Confidence Intervals are in brackets.

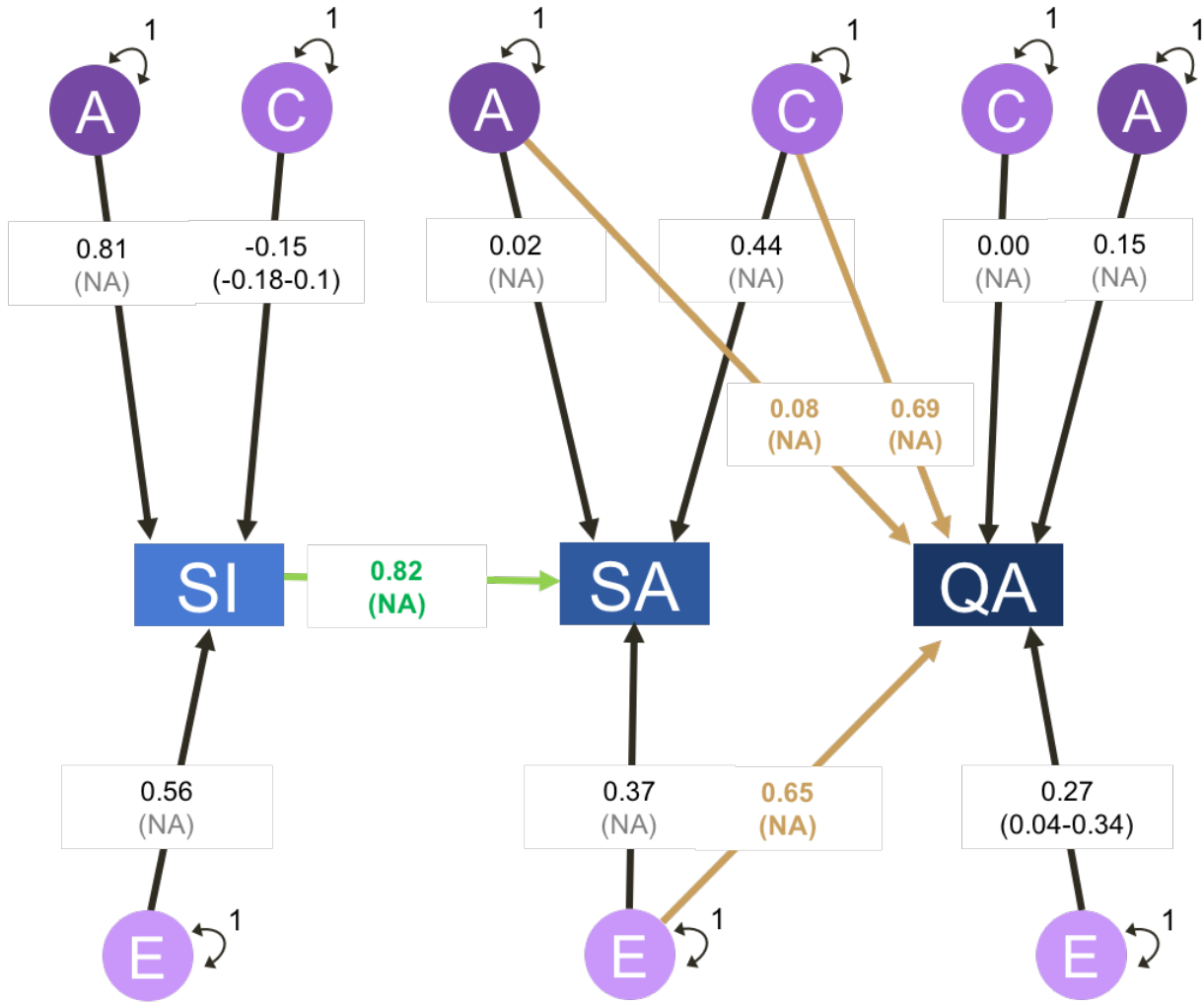


Figure 11. The Best Fitting Model and Path Estimates for Young Adults (age 22-32): VTSABD/TSA. Note: 95% Confidence Intervals are in brackets.

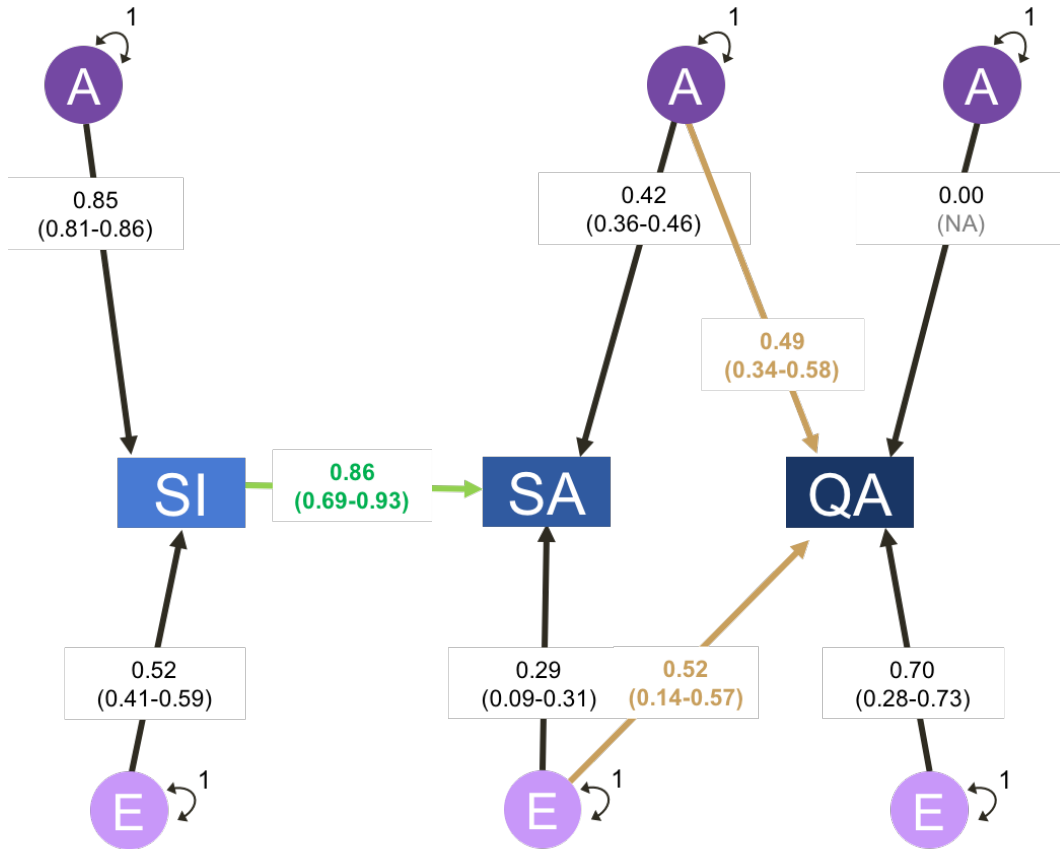


Figure 12. The Best Fitting Model and Path Estimates for All Participants: Add Health.
 Note: 95% Confidence Intervals are in brackets.

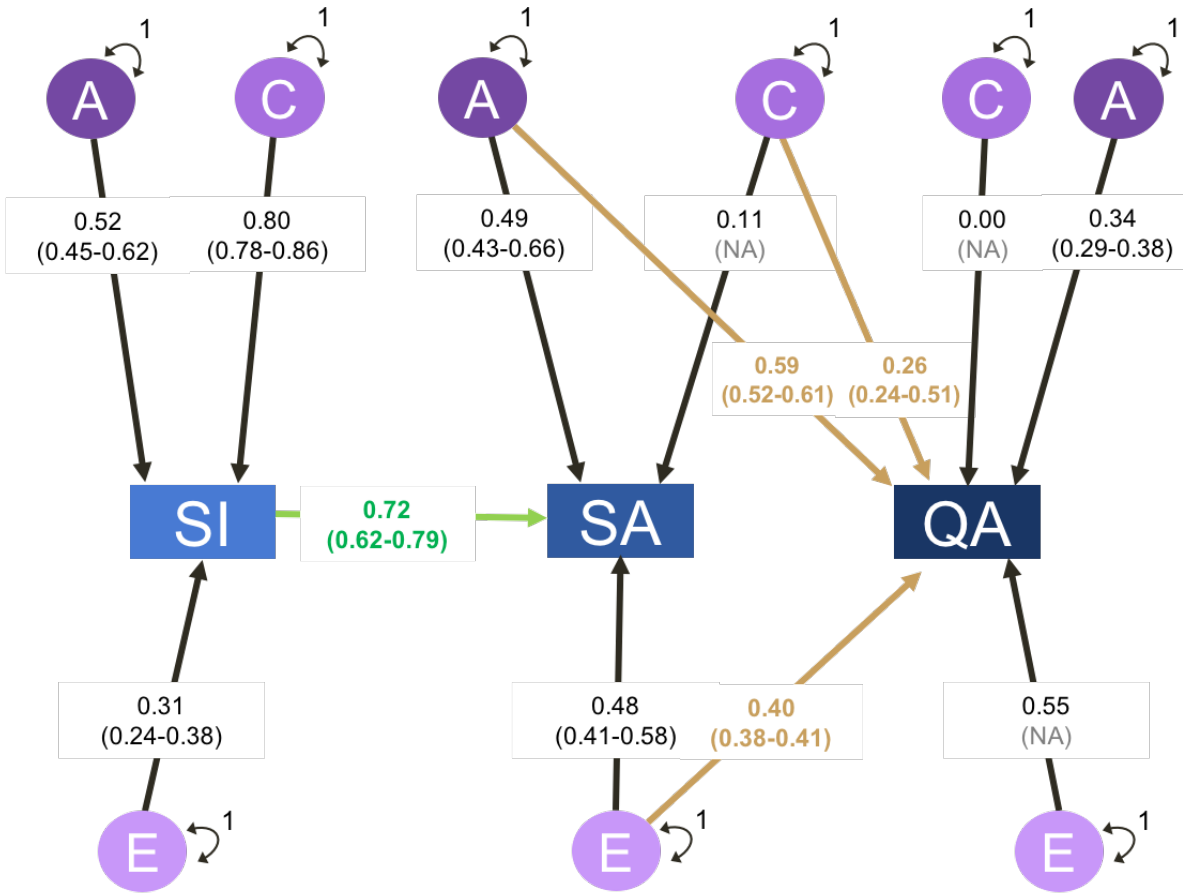


Figure 13. The Best Fitting Model and Path Estimates for All Participants: MASATS. Note: 95% Confidence Intervals are in brackets.

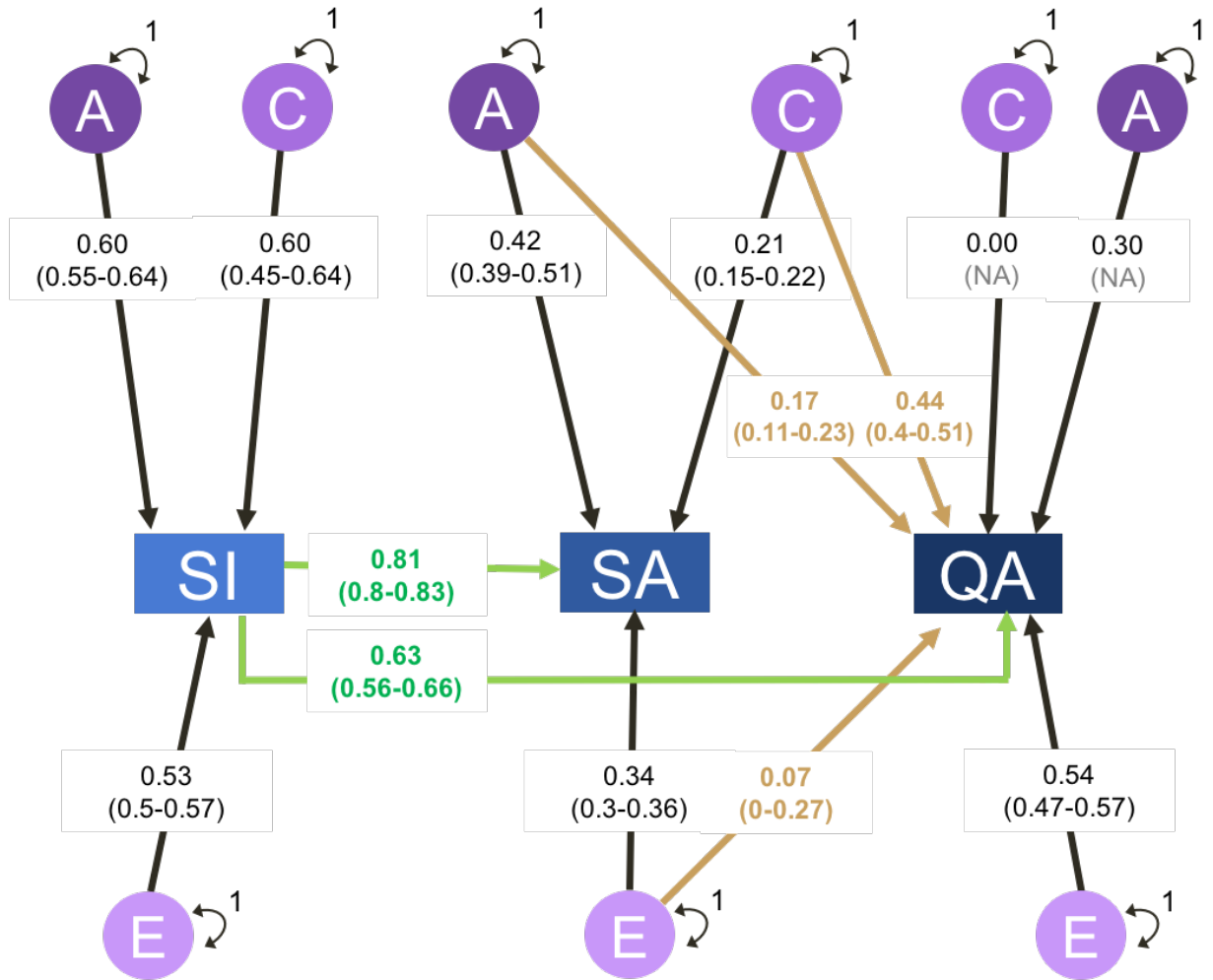


Figure 14. The Best Fitting Model and Path Estimates for All participants: VTSABD/TSA.
 Note: 95% Confidence Intervals are in brackets.

Table 13 Variance Component Estimates from Best Fitting Models

	Add Health				MASATS			VTSABD/TSA	
	12-15	16-19	22-32	All	12-15	16-19	All	22-32	All
Smoking Initiation									
Total A variance	79%	70%	73%	73%	0%	86%	27%	66%	36%
Total C variance	0%	0%	0%	0%	85%	0%	63%	2%	36%
Total E variance	21%	30%	27%	27%	15%	14%	10%	32%	28%
Smoking Amount									
Total A variance	72%	75%	71%	72%	0%	65%	38%	44%	41%
A explained by SI	64%	74%	74%	75%	0%	70%	37%	100%	57%
A unique to SA	36%	26%	26%	25%	0%	30%	63%	0%	43%
Total C variance	0%	0%	0%	0%	69%	0%	34%	21%	28%
C explained by SI	0%	0%	0%	0%	74%	0%	97%	7%	84%
C unique to SA	0%	0%	0%	0%	26%	0%	3%	93%	16%
Total E variance	28%	25%	29%	28%	31%	35%	29%	35%	30%
E explained by SI	42%	95%	68%	71%	30%	21%	18%	61%	61%
E unique to SA	58%	5%	32%	29%	70%	79%	82%	39%	39%
Quit Attempt									
Total A variance	20%	25%	24%	24%	0%	61%	47%	3%	26%
A explained by SI	0%	0%	0%	0%	0%	0%	0%	0%	54%
A explained by SA	85%	100%	100%	100%	0%	65%	75%	23%	11%
A unique to QA	15%	0%	0%	0%	0%	35%	25%	77%	35%
Total C variance	0%	0%	0%	0%	28%	0%	7%	48%	33%
C explained by SI	0%	0%	0%	0%	0%	0%	0%	0%	42%
C explained by SA	0%	0%	0%	0%	55%	0%	100%	100%	58%
C unique to QA	0%	0%	0%	0%	45%	0%	0%	0%	0%
Total E variance	80%	77%	76%	76%	72%	39%	46%	49%	41%
E explained by SI	0%	0%	0%	0%	0%	0%	0%	0%	27%
E explained by SA	49%	57%	26%	35%	70%	49%	34%	85%	1%
E unique to QA	51%	43%	74%	65%	30%	51%	66%	15%	72%

Note: ADH: Add Health, ABD: VTSABD/TSA

Discussion

The goal of this exploratory study was to estimate the role of genetic and environmental contributions specific to smoking quit attempts while accounting for and estimating the genetic and environmental factors shared between quit attempts and smoking initiation as well as smoking amount. This is the first study to investigate the degree and nature of genetic and environmental influences on quit attempts in adolescents and young adults. Currently, most studies focus on smoking abstinence (i.e., current vs. former smoker) particularly in older adults. Encouraging tobacco reduction after initiation is an important public health goal particularly during these important developmental stages. A quit attempt is the first step in the process of smoking cessation and as such it is important to understand the factors involved in this behavior. Therefore, these results are expected to guide future development of public health policies related to smoking cessation and quit attempts specifically. Given the relatively young age of participants, the sample sizes associated with these studies were relatively small and additional replication is suggested. Nevertheless, results highlighted several interesting features of the etiology of smoking quit attempts during adolescence and young adulthood, these include: (1) significant genetic influences on quit attempt, (2) no significant sex differences influencing quit attempts, and (3) there was no significant overlap between QA and SI. These results were consistent across adolescents and young adult.

First Aim: Estimating A, C, and E for QA

This study observed significant A influences specific for lifetime QA across all samples ($A = 3-61\%$), after adjusting for A shared with SI and SA. This means that genes were important for making a decision to quit smoking. However, it is unknown which genes are responsible and

how they influence the decision for quitting smoking. There are no gene discovery studies yet for lifetime quit attempt. Although genetic discovery studies have identified 32 genes for smoking abstinence in adults (Wang and Li 2009), it is unclear whether these genes are also involved in quit attempts and in adolescents.

One possible genetic variant that might influence the decision to quit smoking is the genetic variant in the promoter region of the serotonin transporter gene (5-HTTLPR). This genetic variant is a repeat polymorphism with two alleles, a short (*s*) or long (*l*) allele (14- and 16-repeats, respectively) (Nakamura et al. 2000). This variant is important for decision making (Lesch et al. 1996), i.e. those with an *ss* genotype are more likely to make disadvantageous choices (Homberg et al. 2008) and more likely to choose gambling when facing a guaranteed loss, but are risk-averse when they are offered a guaranteed gain (Rosier et al. 2006; Rosier et al. 2009). The variants in 5-HTTLPR appear to affect amygdala reactivity and regulation, which results in this pattern of decision making (Rosier et al. 2009).

In a way, attempting to quit smoking is a gamble, because success is not a guarantee and might cost smokers socially. I hypothesize that young smokers with *ss* genotype are more likely to take a chance on quitting when faced with the health consequences of smoking, especially if they already have a smoking-related disease. However, I also hypothesize that when a negative impact of quit attempt is guaranteed, such as losing a close group of friends, then young smokers with *ss* genotype are more likely to gamble their health risk for smoking and continue their smoking behavior. This is possible because adolescents who have perceived social benefits of smoking are more likely to smoke (HR = 4.29) (Aryal and Bhatta 2015). Additionally, some smokers have unrealistically optimistic perceptions about their chances of avoiding smoking-related diseases, which results in weaker plans to quit smoking (Dillard, McCaul, and Klein

2006). Knowing a smoker's genotype can help fine-tune smoking cessation approaches for the smoker, especially since smoking cessation programs have low success rates. For example, more efforts on communicating the risk of smoking and counteracting false perceptions are needed for smokers with *ss* genotype. Providing new social circles can also help smokers with *ss* genotype, such as forming a group of teenage smokers who want to quit. Further studies are needed to test these hypotheses and to identify genes important for quit attempts in adolescents.

This study also observed an increase in A across age groups within each sample. For instance, the variance due to genetic influences on QA increased slightly from 20% in younger adolescents to 24% in young adults of ADH. This suggests the presence of developmental changes for the genetic and environmental factors for QA, where genes have increasing importance for deciding to quit as age increases. A possible reason is as adolescents grow up, there are environment changes that increasingly allow the expression of their genetic propensities (Haworth et al. 2009). For example, if the 5-HTTLPR genetic variant hypothesis is true, possibly this gene has an increasing chance to influence decision making for QA because adolescents learn more about the risk of smoking as they age. This learning process can come through receiving more advanced science classes as they age, leading to increasing understanding of the biological and chemical processes of how cigarettes harm health. The health consequences of smoking also presents its symptoms more as smokers age. Thus, the awareness of loss of quality of life increases by age, which triggers decision making process, thus allowing the 5-HTTLPR genetic variant to influence the decision to make a quit attempt. Further studies are needed to confirm and examine the association between genetic variants and quit attempts over time.

For environmental factors, C and E were significant and have a moderate to high influence on lifetime QA in adolescents and young adults (C = 7%-48%, E = 39%-80%). Thus,

this study suggests that environmental influences that are shared between the twins, such as parental behaviors and home environments, and the environments that are unique to each pair, for example friends, partner and work environment, are important for attempting to quit smoking. This agrees with a study of adolescent reasons for quitting smoking (Myers and MacPherson 2008). Long-term concerns about smoking, such as not wanting to have smoking-related diseases in old age and not wanting to be smoking when older, predict QA in the three months following baseline (OR=1.8). However, the effect is relatively short-lived because these long-term concerns did not predict QA over 6-month follow-up (Myers and MacPherson 2008). Instead, social disapproval predicts QA over the 6-month follow-up period (OR=1.1). In the study, social disapproval includes items about disapproval of smoking behaviors from dating partners, friends, parents, and organizations, such as church and youth groups (Myers and MacPherson 2008). It would be interesting to explore these items and other possible reasons for quitting in this study's participants to validate these findings.

Second Aim: Testing sex differences for QA

Different from our hypothesis, we did not observe sex differences for the nature and magnitude of the genetic and environmental influences for QA in this study. The same genetic and environmental factors influenced QA males and females and they affected males and females equally. The thresholds for smoking behaviors could not be equated between sexes in the best fitting models. This might be due to sex differences in the prevalence for SI and SA, not QA (Tables 3-5). This finding suggests that we can use the same public health approaches to promote QA in males and females.

Third Aim: The relationship between SI, SA, and QA

This study demonstrated separate liabilities between QA and SI. This was consistently observed across most data groups. This study suggests that public health policies to prevent SI in Add Health and MASATS might not be effective for promoting QA. However, there are new smoking preventive measures in recent years which are effective for both stages, for example, warning labels. At the time when these twins were interviewed, around the 1990s, warning labels for cigarettes were in form of text. These texts tended to be vague, such as “Smoking can be harmful to your health” (Hammond et al. 2006). There are recent development of graphic warning labels which were first implemented in Canada in 2001 (Jung 2016). These graphic labels are more effective than text labels to prevent SI (Hammond et al. 2006) and increases smoking cessation motivation (Jung 2016). Therefore, an updated study of twins who have been exposed to these types of public policies can potentially show a relationship between SI and QA. Also, a longitudinal study to assess the relationship between factors for SI and QA over time can be useful to monitor the effectiveness of public policies for smoking behaviors.

For VTSABD/TSA participants, the genetic and environmental factors for QA were shared partially with SI. This suggests the need for accounting for SI when identifying the genetic factors for QA. This finding also suggests that smoking preventive measures can help adolescents of VTSABD/TSA make a quit attempt. Identifying the genetic and environmental factors that are shared between SI and QA would help develop effective public health policies for VTSABD/TSA participants.

There were some relationships between the genetic and environmental factors for QA and SA observed in this study. In some data groups, the genetic and environmental factors for QA were completely shared with SA, such as in older Add Health participants. This suggests shared

liabilities between SA and QA. Thus, genetic and environmental factors for SA needed to be accounted for when determining the genetic and environmental factors for QA. Some studies have shown some relationship between SA with SC. For example, SA has been shown to predict successful QA, where those who smoke less were more likely to achieve abstinence (Razavi et al. 1999; Dale et al. 2001; Breslau and Johnson 2000). In addition, there are 10 genes that have been reported for both SA and successful QA (Wang and Li 2009). However, these genes were studied in each stage separately. Further studies are needed to investigate how these genes or other risk factors influence both ND/SA and QA.

Sample differences

There were significant sample differences for QA. First, there were sample differences in QA prevalence ($p < 0.0001$). Second, there were sample differences for the genetic and environmental factor estimates for QA. This might be due to sample size differences. This might also be due to differences in social acceptance for smoking. Add Health is a nationwide study, while MASATS and VTSABD/TSA are from the Mid-Atlantic states. In the US, laws regarding smoking bans are state-specific. For example, in 2013, Kentucky, West Virginia, and Mississippi had no statewide smoking bans for workplace, restaurants, nor bars. Consequently, they had the three highest smoking rates (McCarthy 2015), which might contribute to the sample differences in prevalence. This lack of statewide smoking ban also promoted smoking, potentially creating an environment that limits the influence of genes for quit attempt, thus, leading to sample differences in the genetic and environmental factor estimates for QA.

Limitations

Some limitations of this study are (1) sample sizes, (2) potential recall bias, and (3) limited generalizability of the findings. First, less than half of the study participants have tried to quit smoking and the participants were further divided into age groups. This may reduce the study's power to correctly estimate A/C/E and sex differences, especially in data groups from MASATS. The twin correlation estimates for some of the smoking measures had very large standard errors. Thus these findings should be treated with caution until replicated independently. Additionally, combining the participants in one large analysis, while accounting for age and sample differences, might overcome this. Second, this study relied on retrospective self-report information, which can potentially introduce bias on the phenotype measures. However, the bias might be minimal in multi-wave studies, such as Add Health and VTSABD/TSA. Also, the participants were asked about their smoking behavior that happened near their age at interview. So, recall bias may be minimal in this study. Third, this study was done with participants living in the US and the majority of the participants were of European descents. Thus, these findings might not be generalizable for other ethnic groups or other population.

Conclusion

Genetic and environmental factors significantly contributed to attempting to quit smoking in adolescents and young adults. There were no sex differences for the influences of genetic and environmental factors for QA. The factors for QA were separate from SI in most samples. There were apparent age and sample differences for the genetic and environmental influences for QA.

Further studies are needed identify these genetic and environmental factors to develop effective approaches for promoting QA in adolescent and young adults.

Bibliography

Bibliography

- Agrawal, Arpana, Michael T. Lynskey, Manav Kapoor, Kathleen K. Bucholz, Howard J. Edenberg, Marc Schuckit, Andrew Brooks, et al. 2015. "Are Genetic Variants for Tobacco Smoking Associated with Cannabis Involvement?" *Drug and Alcohol Dependence* 150 (May). Elsevier BV: 183–87. doi:10.1016/j.drugalcdep.2015.02.029.
- Akaike, Hirotugu. 1987. "Factor Analysis and AIC." *Psychometrika* 52 (3). Springer Nature: 317–32. doi:10.1007/bf02294359.
- American Psychiatric Association. 2013. "Diagnostic and Statistical Manual of Mental Disorders 5th Edition American Psychiatric Publishing." *Arlington, VA, USA: Doi* 10.
- Amos, Christopher I, Margaret R Spitz, and Paul Cinciripini. 2010. "Chipping Away at the Genetics of Smoking Behavior." *Nature Genetics* 42 (5). Springer Nature: 366–68. doi:10.1038/ng0510-366.
- Aryal, Umesh Raj, and Dharma Nand Bhatta. 2015. "Perceived Benefits and Health Risks of Cigarette Smoking Among Young Adults: Insights from a Cross-Sectional Study." *Tobacco Induced Diseases* 13 (1). Springer Nature. doi:10.1186/s12971-015-0044-9.
- Asfar, Taghrid, Mark Vander Weg, Wasim Maziak, Fadi Hammal, Thomas Eissenberg, and Kenneth D Ward. 2008. "Outcomes and Adherence in Syria's First Smoking Cessation Trial." *American Journal of Health Behavior* 32 (2). PNG Publications: 146–56.
- Bares, Cristina B., Kenneth S. Kendler, and Hermine H. Maes. 2015. "Developmental Changes in Genetic and Shared Environmental Contributions to Smoking Initiation and Subsequent Smoking Quantity in Adolescence and Young Adulthood." *Twin Research and Human Genetics* 18 (05). Cambridge University Press (CUP): 497–506. doi:10.1017/thg.2015.48.
- Behjati, Sam, and Patrick S Tarpey. 2013. "What Is Next Generation Sequencing?" *Archives of Disease in Childhood - Education & Practice Edition* 98 (6). BMJ: 236–38. doi:10.1136/archdischild-2013-304340.
- Belsky, Daniel W., Terrie E. Moffitt, Timothy B. Baker, Andrea K. Biddle, James P. Evans, HonaLee Harrington, Renate Houts, et al. 2013. "Polygenic Risk and the Developmental Progression to Heavy, Persistent Smoking and Nicotine Dependence." *JAMA Psychiatry*

- 70 (5). American Medical Association (AMA): 534.
doi:10.1001/jamapsychiatry.2013.736.
- Bierut, L. J., P. A.F. Madden, N. Breslau, E. O. Johnson, D. Hatsukami, O. F. Pomerleau, G. E. Swan, et al. 2006. "Novel Genes Identified in a High-Density Genome Wide Association Study for Nicotine Dependence." *Human Molecular Genetics* 16 (1). Oxford University Press (OUP): 24–35. doi:10.1093/hmg/ddl441.
- Bierut, Laura Jean, Stephen H. Dinwiddie, Henri Begleiter, Raymond R. Crowe, Victor Hesselbrock, John I. Nurnberger, Bernice Porjesz, Marc A. Schuckit, and Theodore Reich. 1998. "Familial Transmission of Substance Dependence: Alcohol, Marijuana, Cocaine, and Habitual Smoking." *Archives of General Psychiatry* 55 (11). American Medical Association (AMA): 982. doi:10.1001/archpsyc.55.11.982.
- Bjornson, W, C Rand, J E Connett, P Lindgren, M Nides, F Pope, A S Buist, C Hoppe-Ryan, and P O. 1995. "Gender Differences in Smoking Cessation After 3 Years in the Lung Health Study." *American Journal of Public Health* 85 (2). American Public Health Association: 223–30. doi:10.2105/ajph.85.2.223.
- Bloom, A. Joseph, Sarah M. Hartz, Timothy B. Baker, Li-Shiun Chen, Megan E. Piper, Louis Fox, Maribel Martinez, et al. 2014. "Beyond Cigarettes Per Day. a Genome-Wide Association Study of the Biomarker Carbon Monoxide." *Annals of the American Thoracic Society* 11 (7). American Thoracic Society: 1003–10. doi:10.1513/annalsats.201401-010oc.
- Boker, Steven M., Michael C. Neale, Hermine H. Maes, Michael J. Wilde, Michael Spiegel, Timothy R. Brick, Ryne Estabrook, et al. 2016. *OpenMx 2.6.9 User Guide*.
- Borland, Ron, Neville Owen, David Hill, and Penelope Schofield. 1991. "Predicting Attempts and Sustained Cessation of Smoking After the Introduction of Workplace Smoking Bans." *Health Psychology* 10 (5). American Psychological Association (APA): 336–42. doi:10.1037/0278-6133.10.5.336.
- Breslau, Naomi, and Eric O Johnson. 2000. "Predicting Smoking Cessation and Major Depression in Nicotine-Dependent Smokers." *American Journal of Public Health* 90 (7). American Public Health Association: 1122.
- Broms, Ulla, Pamela A.F. Madden, Andrew C. Heath, Michele L. Pergadia, Saul Shiffman, and Jaakko Kaprio. 2007. "The Nicotine Dependence Syndrome Scale in Finnish Smokers." *Drug and Alcohol Dependence* 89 (1). Elsevier BV: 42–51. doi:10.1016/j.drugalcdep.2006.11.017.

- Broms, Ulla, Karri Silventoinen, Eero Lahelma, Markku Koskenvuo, and Jaakko Kaprio. 2004. "Smoking Cessation by Socioeconomic Status and Marital Status: The Contribution of Smoking Behavior and Family Background." *Nicotine & Tobacco Research* 6 (3). Oxford University Press (OUP): 447–55. doi:10.1080/14622200410001696637.
- Broms, Ulla, Karri Silventoinen, Pamela A. F. Madden, Andrew C. Heath, and Jaakko Kaprio. 2006. "Genetic Architecture of Smoking Behavior: A Study of Finnish Adult Twins." *Twin Research and Human Genetics* 9 (01). Cambridge University Press (CUP): 64–72. doi:10.1375/twin.9.1.64.
- Brown, Tamara, Stephen Platt, and Amanda Amos. 2014. "Equity Impact of Population-Level Interventions and Policies to Reduce Smoking in Adults: A Systematic Review." *Drug and Alcohol Dependence* 138 (May). Elsevier BV: 7–16. doi:10.1016/j.drugalcdep.2014.03.001.
- Bush, William S., and Jason H. Moore. 2012. "Chapter 11: Genome-Wide Association Studies." Edited by Fran Lewitter and Maricel Kann. *PLoS Computational Biology* 8 (12). Public Library of Science (PLoS): e1002822. doi:10.1371/journal.pcbi.1002822.
- Centers for Disease Control and Prevention. 2003. "Cigarette Smoking-Attributable Morbidity—United States, 2000." *MMWR. Morbidity and Mortality Weekly Report* 52 (35): 842.
- . 2011. "Decrease in Smoking Prevalence—Minnesota, 1999-2010." *MMWR. Morbidity and Mortality Weekly Report* 60 (5): 138.
- Cherny, Stacey S. 2009. "QTL Methodology in Behavior Genetics." In *Handbook of Behavior Genetics*, 35–45. Springer.
- Dale, Lowell C., Elbert D. Glover, David P.L. Sachs, Darrell R. Schroeder, Kenneth P. Offord, Ivana T. Croghan, and Richard D. Hurt. 2001. "Bupropion for Smoking Cessation." *Chest* 119 (5). Elsevier BV: 1357–64. doi:10.1378/chest.119.5.1357.
- DiClemente, Carlo C., James O. Prochaska, Scott K. Fairhurst, Wayne F. Velicer, Mary M. Velasquez, and Joseph S. Rossi. 1991. "The Process of Smoking Cessation: An Analysis of Precontemplation, Contemplation, and Preparation Stages of Change." *Journal of Consulting and Clinical Psychology* 59 (2). American Psychological Association (APA): 295–304. doi:10.1037/0022-006x.59.2.295.
- DiFranza, Joseph R., Judith A. Savageau, Kenneth Fletcher, Jennifer O'Loughlin, Lori Pbert, Judith K. Ockene, Ann D. McNeill, et al. 2007. "Symptoms of Tobacco Dependence After Brief Intermittent Use." *Archives of Pediatrics & Adolescent Medicine* 161 (7). American Medical Association (AMA): 704. doi:10.1001/archpedi.161.7.704.

- Dillard, Amanda J., Kevin D. McCaul, and William M. P. Klein. 2006. "Unrealistic Optimism in Smokers: Implications for Smoking Myth Endorsement and Self-Protective Motivation." *Journal of Health Communication* 11 (sup001). Informa UK Limited: 93–102. doi:10.1080/10810730600637343.
- Do, Elizabeth K., Elizabeth C. Prom-Wormley, Lindon J. Eaves, Judy L. Silberg, Donna R. Miles, and Hermine H. Maes. 2015. "Genetic and Environmental Influences on Smoking Behavior Across Adolescence and Young Adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-up." *Twin Research and Human Genetics* 18 (01). Cambridge University Press (CUP): 43–51. doi:10.1017/thg.2014.78.
- D'Onofrio, Brian M, Lenn Murrelle, Lindon J Eaves, Michael E McCullough, Jessica L Landis, and Hermine H Maes. 1999. "Adolescent Religiousness and Its Influence on Substance Use: Preliminary Findings from the MidAtlantic School Age Twin Study." *Twin Research (1999)* 2 (2). Cambridge University Press (CUP): 156–68. doi:10.1375/136905299320566022.
- Elders, M Joycelyn, Cheryl L Perry, Michael P Eriksen, and Gary A Giovino. 1994. "The Report of the Surgeon General: Preventing Tobacco Use Among Young People." *American Journal of Public Health* 84 (4). American Public Health Association: 543–47.
- Emory, K., N. Saquib, E. A. Gilpin, and J. P. Pierce. 2010. "The Association Between Home Smoking Restrictions and Youth Smoking Behaviour: A Review." *Tobacco Control* 19 (6). BMJ: 495–506. doi:10.1136/tc.2010.035998.
- Farkas, A. J, E. A Gilpin, J. M Distefan, and J. P Pierce. 1999. "The Effects of Household and Workplace Smoking Restrictions on Quitting Behaviours." *Tobacco Control* 8 (3). BMJ: 261–65. doi:10.1136/tc.8.3.261.
- Fox, John. 2016. *Polycor: Polychoric and Polyserial Correlations*. <https://CRAN.R-project.org/package=polycor>.
- Frazer, Kelly A., Dennis G. Ballinger, David R. Cox, David A. Hinds, Laura L. Stuve, Richard A. Gibbs, John W. Belmont, et al. 2007. "A Second Generation Human Haplotype Map of over 3.1 Million SNPs." *Nature* 449 (7164). Springer Nature: 851–61. doi:10.1038/nature06258.
- Goriounova, N. A., and H. D. Mansvelder. 2012. "Nicotine Exposure During Adolescence Leads to Short- and Long-Term Changes in Spike Timing-Dependent Plasticity in Rat Prefrontal Cortex." *Journal of Neuroscience* 32 (31). Society for Neuroscience: 10484–93. doi:10.1523/jneurosci.5502-11.2012.

- Gruder, C. L., D. R. Trinidad, P. H. Palmer, B. Xie, L. Li, and C. A. Johnson. 2012. "Tobacco Smoking, Quitting, and Relapsing Among Adult Males in Mainland China: The China Seven Cities Study." *Nicotine & Tobacco Research* 15 (1). Oxford University Press (OUP): 223–30. doi:10.1093/ntr/nts116.
- Hamilton, Ann S, Christina N Lessov-Schlaggar, Myles G Cockburn, Jennifer B Unger, Wendy Cozen, and Thomas M Mack. 2006. "Gender Differences in Determinants of Smoking Initiation and Persistence in California Twins." *Cancer Epidemiology Biomarkers & Prevention* 15 (6). American Association for Cancer Research (AACR): 1189–97. doi:10.1158/1055-9965.epi-05-0675.
- Hammond, David, Geoffrey T Fong, Ann McNeill, Ron Borland, and K Michael Cummings. 2006. "Effectiveness of Cigarette Warning Labels in Informing Smokers About the Risks of Smoking: Findings from the International Tobacco Control (ITC) Four Country Survey." *Tobacco Control* 15 (suppl_3). BMJ: iii19–iii25. doi:10.1136/tc.2005.012294.
- Hardin, J., Y. He, H. S. Javitz, J. Wessel, R. E. Krasnow, E. Tildesley, H. Hops, G. E. Swan, and A. W. Bergen. 2009. "Nicotine Withdrawal Sensitivity, Linkage to Chr6q26, and Association of OPRM1 SNPs in the SMOKing in FAMilies (SMOFAM) Sample." *Cancer Epidemiology Biomarkers & Prevention* 18 (12). American Association for Cancer Research (AACR): 3399–3406. doi:10.1158/1055-9965.epi-09-0960.
- Harris, Kathleen Mullan, Carolyn Tucker Halpern, Andrew Smolen, and Brett C. Haberstick. 2006. "The National Longitudinal Study of Adolescent Health (Add Health) Twin Data." *Twin Research and Human Genetics* 9 (06). Cambridge University Press (CUP): 988–97. doi:10.1375/twin.9.6.988.
- Hatziandreu, E. J., J. P. Pierce, M. Lefkopoulou, M. C. Fiore, S. L. Mills, T. E. Novotny, G. A. Giovino, and R. M. Davis. 1990. "Quitting Smoking in the United States in 1986." *JNCI Journal of the National Cancer Institute* 82 (17). Oxford University Press (OUP): 1402–6. doi:10.1093/jnci/82.17.1402.
- Haworth, C M A, M J Wright, M Luciano, N G Martin, E J C de Geus, C E M van Beijsterveldt, M Bartels, et al. 2009. "The Heritability of General Cognitive Ability Increases Linearly from Childhood to Young Adulthood." *Molecular Psychiatry* 15 (11). Springer Nature: 1112–20. doi:10.1038/mp.2009.55.
- Healthy People 2020. 2016. "Tobacco Use." <https://www.healthypeople.gov/2020/topics-objectives/topic/tobacco-use> [Online; accessed 6-February-2017].

- Heath, A.C., and N.G. Martin. 1993. "Genetic Models for the Natural History of Smoking: Evidence for a Genetic Influence on Smoking Persistence." *Addictive Behaviors* 18 (1). Elsevier BV: 19–34. doi:10.1016/0306-4603(93)90005-t.
- Heath, Andrew C, Pamela AF Madden, and Nicholas G Martin. 1998. "Statistical Methods in Genetic Research on Smoking." *Statistical Methods in Medical Research* 7 (2). SAGE Publications: 165–86. doi:10.1177/096228029800700205.
- Heath, Andrew C., Nicholas G. Martin, Michael T. Lynskey, Alexandre A. Todorov, and Pamela A. F. Madden. 2002. "Estimating Two-Stage Models for Genetic Influences on Alcohol, Tobacco or Drug Use Initiation and Dependence Vulnerability in Twin and Family Data." *Twin Research* 5 (2). Cambridge University Press (CUP): 113–24. doi:10.1375/1369052022983.
- Heatherton, Todd F., Lynn T. Kozlowski, Richard C. Frecker, William Rickert, and Jack Robinson. 1989. "Measuring the Heaviness of Smoking: Using Self-Reported Time to the First Cigarette of the Day and Number of Cigarettes Smoked Per Day." *Addiction* 84 (7). Wiley-Blackwell: 791–800. doi:10.1111/j.1360-0443.1989.tb03059.x.
- Hewitt, John K., Michael Rutter, Emily Simonoff, Andrew Pickles, Rolf Loeber, Andrew C. Heath, Chandra A. Reynolds, et al. 1997. "Genetics and Developmental Psychopathology: 1. Phenotypic Assessment in the Virginia Twin Study of Adolescent Behavioral Development." *Journal of Child Psychology and Psychiatry* 38 (8). Wiley-Blackwell: 943–63. doi:10.1111/j.1469-7610.1997.tb01613.x.
- Hoft, Nicole R, Robin P Corley, Matthew B McQueen, Isabel R Schlaepfer, David Huizinga, and Marissa A Ehringer. 2008. "Genetic Association of the CHRNA6 and CHRNB3 Genes with Tobacco Dependence in a Nationally Representative Sample." *Neuropsychopharmacology* 34 (3). Springer Nature: 698–706. doi:10.1038/npp.2008.122.
- Homberg, Judith R., Ruud van den Bos, Esther den Heijer, Remco Suer, and Edwin Cuppen. 2008. "Serotonin Transporter Dosage Modulates Long-Term Decision-Making in Rat and Human." *Neuropharmacology* 55 (1). Elsevier BV: 80–84. doi:10.1016/j.neuropharm.2008.04.016.
- Husten, C. G. 2009. "How Should We Define Light or Intermittent Smoking? Does It Matter?" *Nicotine & Tobacco Research* 11 (2). Oxford University Press (OUP): 111–21. doi:10.1093/ntr/ntp010.
- Hymowitz, N, K. Cummings, A Hyland, W. Lynn, T. Pechacek, and T. Hartwell. 1997. "Predictors of Smoking Cessation in a Cohort of Adult Smokers Followed for Five

- Years.” *Tobacco Control* 6 (Supplement 2). BMJ: S57–S62.
doi:10.1136/tc.6.suppl_2.s57.
- Jha, Prabhat, and Richard Peto. 2014. “Global Effects of Smoking, of Quitting, and of Taxing Tobacco.” *New England Journal of Medicine* 370 (1). New England Journal of Medicine (NEJM/MMS): 60–68. doi:10.1056/nejmra1308383.
- Jha, Prabhat, Chinthanie Ramasundarahettige, Victoria Landsman, Brian Rostron, Michael Thun, Robert N. Anderson, Tim McAfee, and Richard Peto. 2013. “21st-Century Hazards of Smoking and Benefits of Cessation in the United States.” *New England Journal of Medicine* 368 (4). New England Journal of Medicine (NEJM/MMS): 341–50.
doi:10.1056/nejmsa1211128.
- Johnston, Lloyd D, Patrick M O’malley, Jerald G Bachman, and John E Schulenberg. 2011. “Monitoring the Future National Survey Results on Drug Use, 1975-2010. Volume Ii, College Students & Adults Ages 19-50.” *Institute for Social Research*. ERIC.
- Joossens, L. 2006. “The Tobacco Control Scale: A New Scale to Measure Country Activity.” *Tobacco Control* 15 (3). BMJ: 247–53. doi:10.1136/tc.2005.015347.
- Jung, Minsoo. 2016. “Implications of Graphic Cigarette Warning Labels on Smoking Behavior: An International Perspective.” *Journal of Cancer Prevention* 21 (1). Korean Society of Cancer Prevention: 21–25. doi:10.15430/jcp.2016.21.1.21.
- Karpinski, Julie P, Erin M Timpe, and Lisa Lubsch. 2010. “Smoking Cessation Treatment for Adolescents.” *The Journal of Pediatric Pharmacology and Therapeutics* 15 (4). Pediatric Pharmacy Advocacy Group: 249–63.
- Kendler, K. S., M. C. Neale, P. Sullivan, L. A. Corey, C. O. Gardner, and C. A. Prescott. 1999. “A Population-Based Twin Study in Women of Smoking Initiation and Nicotine Dependence.” *Psychological Medicine* 29 (2). Cambridge University Press (CUP): 299–308. doi:10.1017/s0033291798008022.
- Keskitalo, K., U. Broms, M. Heliövaara, S. Ripatti, I. Surakka, M. Perola, J. Pitkaniemi, L. Peltonen, A. Aromaa, and J. Kaprio. 2009. “Association of Serum Cotinine Level with a Cluster of Three Nicotinic Acetylcholine Receptor Genes (CHRNA3/CHRNA5/CHRNA4) on Chromosome 15.” *Human Molecular Genetics* 18 (20). Oxford University Press (OUP): 4007–12. doi:10.1093/hmg/ddp322.
- Khuder, Sadik A, Hari H Dayal, and Anand B Mutgi. 1999. “Age at Smoking Onset and Its Effect on Smoking Cessation.” *Addictive Behaviors* 24 (5). Elsevier BV: 673–77.
doi:10.1016/s0306-4603(98)00113-0.

- Kim, Young Jin, and Juyoung Lee, Bong-Jo Kim, and Taesung Park. 2015. "A New Strategy for Enhancing Imputation Quality of Rare Variants from Next-Generation Sequencing Data via Combining SNP and Exome Chip Data." *BMC Genomics* 16 (1). Springer Nature. doi:10.1186/s12864-015-2192-y.
- Lee, Chung-won, and Jennifer Kahende. 2007. "Factors Associated with Successful Smoking Cessation in the United States, 2000." *American Journal of Public Health* 97 (8). American Public Health Association: 1503–9. doi:10.2105/ajph.2005.083527.
- Lesch, K.-P., D. Bengel, A. Heils, S. Z. Sabol, B. D. Greenberg, S. Petri, J. Benjamin, C. R. Muller, D. H. Hamer, and D. L. Murphy. 1996. "Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region." *Science* 274 (5292). American Association for the Advancement of Science (AAAS): 1527–31. doi:10.1126/science.274.5292.1527.
- Li, Ming D., Rong Cheng, Jennie Z. Ma, and Gary E. Swan. 2003. "A Meta-Analysis of Estimated Genetic and Environmental Effects on Smoking Behavior in Male and Female Adult Twins." *Addiction* 98 (1). Wiley-Blackwell: 23–31. doi:10.1046/j.1360-0443.2003.00295.x.
- Li, Yun, Cristen Willer, Serena Sanna, and Gonçalo Abecasis. 2009. "Genotype Imputation." *Annual Review of Genomics and Human Genetics* 10 (1). Annual Reviews: 387–406. doi:10.1146/annurev.genom.9.081307.164242.
- Lichtenstein, Edward, and Robin J. Mermelstein. 1986. "Some Methodological Cautions in the Use of the Tolerance Questionnaire." *Addictive Behaviors* 11 (4). Elsevier BV: 439–42. doi:10.1016/0306-4603(86)90024-9.
- Liu, Y-Z, Y-F Pei, Y-F Guo, L Wang, X-G Liu, H Yan, D-H Xiong, et al. 2009. "Genome-Wide Association Analyses Suggested a Novel Mechanism for Smoking Behavior Regulated by IL15." *Molecular Psychiatry* 14 (7). Springer Nature: 668–80. doi:10.1038/mp.2009.3.
- Lund, Marianne. 2015. "Social Inequality in Cigarette Consumption, Cigarette Dependence, and Intention to Quit Among Norwegian Smokers." *BioMed Research International* 2015. Hindawi Publishing Corporation: 1–7. doi:10.1155/2015/835080.
- Madden, Pamela A. F., Nancy L. Pedersen, Jaakko Kaprio, Markku J. Koskenvuo, and Nicholas G. Martin. 2004. "The Epidemiology and Genetics of Smoking Initiation and Persistence: Crosscultural Comparisons of Twin Study Results." *Twin Research* 7 (1). Cambridge University Press (CUP): 82–97. doi:10.1375/13690520460741471.

- Madden, Pamela AF, Andrew C Heath, Nancy L Pedersen, Jaakko Kaprio, Markku J Koskenvuo, and Nicholas G Martin. 1999. "The Genetics of Smoking Persistence in Men and Women: A Multicultural Study." *Behavior Genetics* 29 (6). Springer: 423–31. doi:10.1023/A:1021674804714.
- Maes, Hermine H., Elizabeth Prom-Wormley, Lindon J. Eaves, Soo Hyun Rhee, John K. Hewitt, Susan Young, Robin Corley, et al. 2016. "A Genetic Epidemiological Mega Analysis of Smoking Initiation in Adolescents." *Nicotine & Tobacco Research*, November. Oxford University Press (OUP), ntw294. doi:10.1093/ntr/ntw294.
- Maes, Hermine H., Patrick F. Sullivan, Cynthia M. Bulik, Michael C. Neale, Carol A. Prescott, Lindon J. Eaves, and Kenneth S. Kendler. 2004. "A Twin Study of Genetic and Environmental Influences on Tobacco Initiation, Regular Tobacco Use and Nicotine Dependence." *Psychological Medicine* 34 (7). Cambridge University Press (CUP): 1251–61. doi:10.1017/s0033291704002405.
- McCarthy, Justin. 2015. "In Us, Smoking Rate Lowest in Utah, Highest in Kentucky."
- Morley, Katherine I, Michael T Lynskey, Pamela AF Madden, Susan A Treloar, Andrew C Heath, and Nicholas G Martin. 2007. "Exploring the Inter-Relationship of Smoking Age-at-Onset, Cigarette Consumption and Smoking Persistence: Genes or Environment?" *Psychological Medicine* 37 (09). Cambridge University Press (CUP): 1357. doi:10.1017/s0033291707000748.
- Murray, Robert P., Lynn B. Gerald, Paula G. Lindgren, John E. Connett, Cynthia S. Rand, and Nicholas R. Anthonisen. 2000. "Characteristics of Participants Who Stop Smoking and Sustain Abstinence for 1 and 5 Years in the Lung Health Study." *Preventive Medicine* 30 (5). Elsevier BV: 392–400. doi:10.1006/pmed.2000.0642.
- Myers, Mark G., and Laura MacPherson. 2008. "Adolescent Reasons for Quitting Smoking: Initial Psychometric Evaluation." *Psychology of Addictive Behaviors* 22 (1). American Psychological Association (APA): 129–34. doi:10.1037/0893-164x.22.1.129.
- Nakamura, M, S Ueno, A Sano, and H Tanabe. 2000. "The Human Serotonin Transporter Gene Linked Polymorphism (5-HTTLPR) Shows Ten Novel Allelic Variants." *Molecular Psychiatry* 5 (1). Springer Nature: 32–38. doi:10.1038/sj.mp.4000698.
- Neale, MC, and LR Cardon. 1992. "Methodology for Genetic Studies of Twins and Families." Kluwer: Dordrecht.
- Neale, Michael C. 2009. "Biometrical Models in Behavioral Genetics." In *Handbook of Behavior Genetics*, 15–33. Springer.

- Neale, Michael C., Michael D. Hunter, Joshua N. Pritikin, Mahsa Zahery, Timothy R. Brick, Robert M. Kickpatrick, Ryne Estabrook, Timothy C. Bates, Hermine H. Maes, and Steven M. Boker. 2016. "OpenMx 2.0: Extended Structural Equation and Statistical Modeling." *Psychometrika* 81 (2). Psychometric Society: 535–49. doi:10.1007/s11336-014-9435-8.
- Neyman, J., and E. S. Pearson. 1928. "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I." *Biometrika* 20A (1/2). JSTOR: 175. doi:10.2307/2331945.
- Okuyemi, Kolawole, Kari Jo Harris, Monica Scheibmeir, Won Choi, Joshua Powell, and Jasjit Ahluwalia. 2002. "Light Smokers: Issues and Recommendations." *Nicotine & Tobacco Research* 4 (4). Oxford University Press (OUP): 103–12. doi:10.1080/1462220021000032726.
- O'Loughlin, Jennifer, Marie-Pierre Sylvestre, Aurélie Labbe, Nancy C Low, Marie-Hélène Roy-Gagnon, Erika N Dugas, Igor Karp, and James C Engert. 2014. "Genetic Variants and Early Cigarette Smoking and Nicotine Dependence Phenotypes in Adolescents." Edited by Huiping Zhang. *PLoS ONE* 9 (12). Public Library of Science (PLoS): e115716. doi:10.1371/journal.pone.0115716.
- Panday, Saadhna, S. Priscilla Reddy, Robert A.C. Ruiter, Erik Bergström, and Hein de Vries. 2007. "Nicotine Dependence and Withdrawal Symptoms Among Occasional Smokers." *Journal of Adolescent Health* 40 (2). Elsevier BV: 144–50. doi:10.1016/j.jadohealth.2006.09.001.
- Park, Sunhee, Terri E. Weaver, and Daniel Romer. 2010. "Predictors of the Transition from Experimental to Daily Smoking in Late Adolescence and Young Adulthood." *Journal of Drug Education* 40 (2). SAGE Publications: 125–41. doi:10.2190/de.40.2.b.
- Piper, Megan E., Thomas M. Piasecki, E. Belle Federman, Daniel M. Bolt, Stevens S. Smith, Michael C. Fiore, and Timothy B. Baker. 2004. "A Multiple Motives Approach to Tobacco Dependence: The Wisconsin Inventory of Smoking Dependence Motives (WISDM-68)." *Journal of Consulting and Clinical Psychology* 72 (2). American Psychological Association (APA): 139–54. doi:10.1037/0022-006x.72.2.139.
- Plomin, R. 2011. "Commentary: Why Are Children in the Same Family so Different? Non-Shared Environment Three Decades Later." *International Journal of Epidemiology* 40 (3). Oxford University Press (OUP): 582–92. doi:10.1093/ije/dyq144.

- Pritikin, Joshua N., Michael D. Hunter, and Steven M. Boker. 2015. "Modular Open-Source Software for Item Factor Analysis." *Educational and Psychological Measurement* 75 (3): 458–74.
- Prokhorov, Alexander V., Unto E. Pallonen, Joseph L. Fava, Lin Ding, and Raymond Niaura. 1996. "Measuring Nicotine Dependence Among High-Risk Adolescent Smokers." *Addictive Behaviors* 21 (1). Elsevier BV: 117–27. doi:10.1016/0306-4603(96)00048-2.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Razavi, D, H Vandecasteele, C Primo, M Bodo, F Debrier, H Verbist, D Pethica, M Eerdeken, and L Kaufman. 1999. "Maintaining Abstinence from Cigarette Smoking: Effectiveness of Group Counselling and Factors Predicting Outcome." *European Journal of Cancer* 35 (8). Elsevier BV: 1238–47. doi:10.1016/s0959-8049(99)00110-0.
- Revelle, William. 2016. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. <https://CRAN.R-project.org/package=psych>.
- Richmond, Robyn L, Anne Austin, and Ian W Webster. 1988. "Predicting Abstainers in a Smoking Cessation Programme Administered by General Practitioners." *International Journal of Epidemiology* 17 (3). Oxford University Press (OUP): 530–34. doi:10.1093/ije/17.3.530.
- Rodgman, Alan, and Thomas A Perfetti. 2013. *The Chemical Components of Tobacco and Tobacco Smoke*. CRC press.
- Roiser, J. P., B. de Martino, G. C. Y. Tan, D. Kumaran, B. Seymour, N. W. Wood, and R. J. Dolan. 2009. "A Genetically Mediated Bias in Decision Making Driven by Failure of Amygdala Control." *Journal of Neuroscience* 29 (18). Society for Neuroscience: 5985–91. doi:10.1523/jneurosci.0407-09.2009.
- Roiser, Jonathan P., Robert D. Rogers, Lynnette J. Cook, and Barbara J. Sahakian. 2006. "The Effect of Polymorphism at the Serotonin Transporter Gene on Decision-Making, Memory and Executive Function in Ecstasy Users and Controls." *Psychopharmacology* 188 (2). Springer Nature: 213–27. doi:10.1007/s00213-006-0495-z.
- Rojas, Neal L., Joel D. Killen, K. Farish Haydel, and Thomas N. Robinson. 1998. "Nicotine Dependence Among Adolescent Smokers." *Archives of Pediatrics & Adolescent Medicine* 152 (2). American Medical Association (AMA). doi:10.1001/archpedi.152.2.151.

- Rose, Richard J, Ulla Broms, Tellervo Korhonen, Danielle M Dick, and Jaakko Kaprio. 2009. "Genetics of Smoking Behavior." In *Handbook of Behavior Genetics*, 411–32. Springer.
- Saccone, Nancy L., Scott F. Saccone, Anthony L. Hinrichs, Jerry A. Stitzel, Weimin Duan, Michele L. Pergadia, Arpana Agrawal, et al. 2009. "Multiple Distinct Risk Loci for Nicotine Dependence Identified by Dense Coverage of the Complete Family of Nicotinic Receptor Subunit (CHRN) Genes." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 150B (4). Wiley-Blackwell: 453–66. doi:10.1002/ajmg.b.30828.
- Saccone, S. F., A. L. Hinrichs, N. L. Saccone, G. A. Chase, K. Konvicka, P. A.F. Madden, N. Breslau, et al. 2006. "Cholinergic Nicotinic Receptor Genes Implicated in a Nicotine Dependence Association Study Targeting 348 Candidate Genes with 3713 SNPs." *Human Molecular Genetics* 16 (1). Oxford University Press (OUP): 36–49. doi:10.1093/hmg/ddl438.
- Scherrer, Jeffrey F., Hong Xian, Hui Pan, Michele L. Pergadia, Pamela A.F. Madden, Julia D. Grant, Carolyn E. Sartor, Jon Randolph Haber, Theodore Jacob, and Kathleen K. Bucholz. 2012. "Parent, Sibling and Peer Influences on Smoking Initiation, Regular Smoking and Nicotine Dependence. Results from a Genetically Informative Design." *Addictive Behaviors* 37 (3). Elsevier BV: 240–47. doi:10.1016/j.addbeh.2011.10.005.
- Senore, Carlo, Renaldo N. Battista, Stanley H. Shapiro, Nereo Segnan, Antonio Ponti, Stefano Rosso, and Daniela Aimar. 1998. "Predictors of Smoking Cessation Following Physician Counseling." *Preventive Medicine* 27 (3). Elsevier BV: 412–21. doi:10.1006/pmed.1998.0286.
- Shiffman, Saul, and Michael A. Sayette. 2005. "Validation of the Nicotine Dependence Syndrome Scale (NDSS): A Criterion-Group Design Contrasting Chippers and Regular Smokers." *Drug and Alcohol Dependence* 79 (1). Elsevier BV: 45–52. doi:10.1016/j.drugalcdep.2004.12.009.
- Simonoff, Emily. 1997. "The Virginia Twin Study of Adolescent Behavioral Development." *Archives of General Psychiatry* 54 (9). American Medical Association (AMA): 801. doi:10.1001/archpsyc.1997.01830210039004.
- Sims, T. H. 2009. "Tobacco as a Substance of Abuse." *PEDIATRICS* 124 (5). American Academy of Pediatrics (AAP): e1045–e1053. doi:10.1542/peds.2009-2121.
- Stanton, Warren R., John B. Lowe, and Amaya M. Gillespie. 1996. "Adolescents Experiences of Smoking Cessation." *Drug and Alcohol Dependence* 43 (1-2). Elsevier BV: 63–70. doi:10.1016/s0376-8716(97)84351-7.

- Stapleton, John A., Michael A. H. Russell, Colin Feyerabend, Simon M Wiseman, Gunnar Gustavsson, Urbain Sawe, and Debbie Wiseman. 2006. "Dose Effects and Predictors of Outcome in a Randomized Trial of Transdermal Nicotine Patches in General Practice." *Addiction* 90 (1). Wiley-Blackwell: 31–42. doi:10.1046/j.1360-0443.1995.901316.x.
- Sussman, S. 2003. "Effects of Sixty Six Adolescent Tobacco Use Cessation Trials and Seventeen Prospective Studies of Self-Initiated Quitting." *Tobacco Induced Diseases* 1 (1). Springer Nature: 35. doi:10.1186/1617-9625-1-1-35.
- Tobacco and Genetics Consortium. 2010. "Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior." *Nature Genetics* 42 (5). Springer Nature: 441–47. doi:10.1038/ng.571.
- Tonnesen, P, P Paoletti, G Gustavsson, M.a Russell, R Saracci, A Gulsvik, B Rijcken, U Sawe, and members of the Steering Committee. 1999. "Higher Dosage Nicotine Patches Increase One-Year Smoking Cessation Rates: Results from the European CEASE Trial." *European Respiratory Journal* 13 (2). European Respiratory Society (ERS): 238. doi:10.1034/j.1399-3003.1999.13b04.x.
- True, William R., Andrew C. Heath, Jeffrey F. Scherrer, Brian Waterman, Jack Goldberg, Nong Lin, Seth A. Eisen, Michael J. Lyons, and Ming T. Tsuang. 1997. "Genetic and Environmental Contributions to Smoking." *Addiction* 92 (10). Wiley-Blackwell: 1277–87. doi:10.1111/j.1360-0443.1997.tb02847.x.
- Turner, Lindsey, Robin Mermelstein, and Brian Flay. 2004. "Individual and Contextual Influences on Adolescent Smoking." *Annals of the New York Academy of Sciences* 1021 (1). Wiley-Blackwell: 175–97. doi:10.1196/annals.1308.023.
- Tyas, S. L, and L. L Pederson. 1998. "Psychosocial Factors Related to Adolescent Smoking: A Critical Review of the Literature." *Tobacco Control* 7 (4). BMJ: 409–20. doi:10.1136/tc.7.4.409.
- Uhl, George R, Qing-Rong Liu, Tomas Drgon, Catherine Johnson, Donna Walther, and Jed E Rose. 2007. "Molecular Genetics of Nicotine Dependence and Abstinence: Whole Genome Association Using 520,000 SNPs." *BMC Genetics* 8 (1). Springer Nature: 10. doi:10.1186/1471-2156-8-10.
- Uhl, George R., Qing-Rong Liu, Tomas Drgon, Catherine Johnson, Donna Walther, Jed E. Rose, Sean P. David, Ray Niaura, and Caryn Lerman. 2008. "Molecular Genetics of Successful Smoking Cessation." *Archives of General Psychiatry* 65 (6). American Medical Association (AMA): 683. doi:10.1001/archpsyc.65.6.683.

- US Department of Health and Human Services. 1990. "The Surgeon General's 1990 Report on the Health Benefits of Smoking Cessation." *Rockville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*, 89–8411.
- . 2004. "The Health Consequences of Smoking: A Report of the Surgeon General." *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*.
- . 2010. "How Tobacco Smoke Causes Disease—The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General." *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*.
- . 2014. "The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General." *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*.
- Vink, J M, A L Beem, D Posthuma, M C Neale, G Willemsen, K S Kendler, P E Slagboom, and D I Boomsma. 2004. "Linkage Analysis of Smoking Initiation and Quantity in Dutch Sibling Pairs." *The Pharmacogenomics Journal* 4 (4). Springer Nature: 274–82. doi:10.1038/sj.tpj.6500255.
- Vink, Jacqueline M., Gonneke Willemsen, and Dorret I. Boomsma. 2005. "Heritability of Smoking Initiation and Nicotine Dependence." *Behavior Genetics* 35 (4). Springer Nature: 397–406. doi:10.1007/s10519-004-1327-8.
- Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57 (2). JSTOR: 307. doi:10.2307/1912557.
- Wang, Ju, and Ming D Li. 2009. "Common and Unique Biological Pathways Associated with Smoking Initiation/Progression, Nicotine Dependence, and Smoking Cessation." *Neuropsychopharmacology* 35 (3). Springer Nature: 702–19. doi:10.1038/npp.2009.178.
- Wiener, R. Constance, Alcinda K. Trickett Shockey, and Susan K. Morgan. 2016. "Adolescent Light Cigarette Smoking Patterns and Adult Cigarette Smoking." *Advances in Epidemiology* 2016. Hindawi Publishing Corporation: 1–7. doi:10.1155/2016/9587340.

- Wigand, JS. 2006. "Additives, Cigarette Design and Tobacco Product Regulation. a Report to: World Health Organization, Tobacco Free Initiative, Tobacco Product Regulation Group, 28 June–2 July 2006, Kobe, Japan." In, 1–45.
- Wilkinson, Anna V., Matthew B. Schabath, Alexander V. Prokhorov, and Margaret R. Spitz. 2007. "Age-Related Differences in Factors Associated with Smoking Initiation." *Cancer Causes & Control* 18 (6). Springer Nature: 635–44. doi:10.1007/s10552-007-9008-6.
- Williams, Larry J., and Patricia J. Holahan. 1994. "Parsimony-Based Fit Indices for Multiple-Indicator Models: Do They Work?" *Structural Equation Modeling: A Multidisciplinary Journal* 1 (2). Informa UK Limited: 161–89. doi:10.1080/10705519409539970.
- Wilson, David, Melanie Wakefield, Neville Owen, and Lyn Roberts. 1992. "Characteristics of Heavy Smokers." *Preventive Medicine* 21 (3). Elsevier BV: 311–19. doi:10.1016/0091-7435(92)90030-1.
- World Health Organization. 2011. *WHO Report on the Global Tobacco Epidemic, 2011: Warning About the Dangers of Tobacco*. World Health Organization.
- Wray, Naomi, and P Visscher. 2008. "Estimating Trait Heritability." *Nature Education* 1 (1): 29.
- Xian, H., J. F. Scherrer, P. A. F. Madden, M. J. Lyons, M. Tsuang, W. R. True, and S. A. Eisen. 2003. "The Heritability of Failed Smoking Cessation and Nicotine Withdrawal in Twins Who Smoked and Attempted to Quit." *Nicotine & Tobacco Research* 5 (2). Oxford University Press (OUP): 245–54. doi:10.1080/1462220031000073667.
- Zhong, Yan, Steven G. Carmella, Pramod Upadhyaya, J. Bradley Hochalter, Diane Rauch, Andrew Oliver, Joni Jensen, et al. 2011. "Immediate Consequences of Cigarette Smoking: Rapid Formation of Polycyclic Aromatic Hydrocarbon Diol Epoxides." *Chemical Research in Toxicology* 24 (2). American Chemical Society (ACS): 246–52. doi:10.1021/tx100345x.
- Zhu, Shu-Hong, Ted Melcer, Jichao Sun, Bradley Rosbrook, and John P Pierce. 2000. "Smoking Cessation with and Without Assistance." *American Journal of Preventive Medicine* 18 (4). Elsevier BV: 305–11. doi:10.1016/s0749-3797(00)00124-0.

Appendix

Appendix I R script for building datasets (ABD: VTSABD/TSA; ADH: Add Health; NCT: MASATS)

```
# ----- Smoking Abstinence -----  
  
# ----- Packages -----  
  
library('car')  
library('gmodels')  
library(MASS)  
library("polycor")  
  
# ---- Separate Datasets ----  
  
# 1. ABD  
# 2. ADH  
# 3. NCT  
  
# ---- ABD ----  
  
ABD <- read.table(file="ABDquit.csv", header = TRUE, sep = ",")  
  
# cigs/day. artobnc = 0, then cpd = 0. if 1-10, cpd = 1, 11- 20, cpd = 2, 21-30 cpd 3 more than 31, cpd 4  
# wave 6 is TSA  
  
ABD$maxcigxdayinf <- ifelse (is.na(ABD$cigxdayw1) & is.na(ABD$cigxdayw2) & is.na(ABD$cigxdayw3) &  
                           is.na(ABD$cigxdayw4) & is.na(ABD$cigxdayw6), 0.1, 0)  
  
ABD$maxcigxday1 <- as.numeric(apply(ABD[c("cigxdayw1", "cigxdayw2", "cigxdayw3", "cigxdayw4",  
"cigxdayw6", "maxcigxdayinf")], 1, max, na.rm=TRUE))  
  
ABD$maxcigxday <- as.numeric(1*(recode(ABD$maxcigxday1, "0=0; 1=1; 2=2; 3=3; 0.1=NA")))  
  
ABDbuild <- ABD  
  
# ---- ABD: recode sex for zygoty = 5 -----  
  
# make male twins as twin 1, and female twins as twin 2  
  
ABDbuild$IndId_int <- ifelse(ABDbuild$Zygroup == 5 & ABDbuild$sex == 1, 1,  
                           ifelse (ABDbuild$Zygroup == 5 & ABDbuild$sex == 2, 2, ABDbuild$IndId))  
  
# ----- ABD: Remove missing tobever, sex, and zygoty -----  
  
ABDbuild <- ABDbuild[-which(is.na(ABDbuild$tobever)),]  
table(ABDbuild$Zygroup, useNA = 'always') # 23 still missing
```

```

ABDbuild <- ABDbuild[-which(is.na(ABDbuild$zygroup)),]
table(ABDbuild$sex, useNA = 'always') # sex no longer missing
dim(ABD)[1]- dim(ABDbuild)[1] # 25 individuals removed
# ----- ABD: recode conflicting maxcigxday -----
table(ABDbuild$tobever, ABDbuild$maxcigxday, useNA = 'always') # 1 nonsmoker with maxcigxday = 1
ABDbuild$tobever[ABDbuild$tobever == 0 & ABDbuild$maxcigxday > 0] <- 1
table(ABDbuild$tobever, ABDbuild$maxcigxday, useNA = 'always')
# set nonsmokers to have missing value
ABDbuild$maxcigxdayCCC <- ifelse(ABDbuild$tobever==0, NA, ABDbuild$maxcigxday)
ABDbuild$maxcigxdayCCC <- ifelse(ABDbuild$maxcigxdayCCC == 0, 0,
                                ifelse(ABDbuild$maxcigxdayCCC == 1, 1, 2))
table(ABDbuild$tobever, ABDbuild$maxcigxdayCCC, useNA = 'always')
# ----- ABD: recode quitever -----
table(ABDbuild$tobever, ABDbuild$quitty, useNA = 'always') # non conflicting
ABDbuild$quiteverCCC <- ifelse(ABDbuild$tobever==0, NA, ABDbuild$quitty)
table(ABDbuild$tobever, ABDbuild$quiteverCCC, useNA = 'always')
# ---- ABD: final dataset ----
ABDbuildfull <- ABDbuild
ABDbuild <- subset(ABDbuildfull, select = c("FAMNO", "IndId", "IndId_int", "StudyNum", "StudyName",
                                           "zygroup", "sex", "ethnicity", "w6age",
                                           "tobever",
                                           "maxcigxdayCCC", "quiteverCCC"))
# ---- Subset only the TSA (adults) ----
ABDTSA <- ABDbuild[-which(is.na(ABDbuild$w6age)),]
write.csv(ABDbuild, "ABDbuild.csv")
write.csv(ABDTSA, "ABDGroupTSA.csv")
# ---- ADH ----
ADH <- read.csv("ADHquits.csv", header=T)
# H3TO7 cigs per day past 30
# CHECK cigxdayw3: IF MISSING and tobever=1, then have cigxdayw3 = 0

```

```

ADH$cigxdayw3 <- ifelse(is.na(ADH$cigxdayw3)&ADH$H3TO7==0, 0, ADH$cigxdayw3)

ADH$maxcigxdayinf <- ifelse (is.na(ADH$cigxdayw1) & is.na(ADH$cigxdayw2) & is.na(ADH$cigxdayw3) &
                             is.na(ADH$cigxdayw4), 0.1, 0)

ADH$maxcigxday1 <- as.numeric(apply(ADH[c("cigxdayw1", "cigxdayw2", "cigxdayw3", "cigxdayw4",
"maxcigxdayinf")], 1, max, na.rm=TRUE))

ADH$maxcigxday <- as.numeric(1*(recode(ADH$maxcigxday1, "0=0; 1=1; 2=2; 3=3; 0.1=NA")))

ADHbuild <- ADH

# ---- ADH: recode sex for zygosity = 5 -----

# make male twins as twin 1, and female twins as twin 2

ADHbuild$IndId_int <- ifelse(ADHbuild$zygroup == 5 & !is.na(ADHbuild$sex) & ADHbuild$sex == 1, 1,
                             ifelse (ADHbuild$zygroup == 5 & !is.na(ADHbuild$sex) & ADHbuild$sex == 2, 2,
ADHbuild$IndId))

# ----- ADH: check quitever -----

ADHbuild$quitever1<-recode(ADHbuild$H1TO8, "0=0; 1=1; 6=NA; 7=NA; 8=NA ; 9=NA")
ADHbuild$quitever2<-recode(ADHbuild$H2TO9, "0=0; 1=1; 6=NA; 7=NA; 8=NA; 9=NA")
ADHbuild$quitever3<-recode(ADHbuild$H3TO11, "0=0; 1=1; 6=NA; 7=NA; 8=NA; 9=NA")
ADHbuild$quitever4<-recode(ADHbuild$H4TO27, "0=0; 1=1; 6=NA; 7=NA; 8=NA; 9=NA")

ADHbuild$quitever <- ifelse(is.na(ADHbuild$quitever1)&
                             is.na(ADHbuild$quitever2)&
                             is.na(ADHbuild$quitever3)&
                             is.na(ADHbuild$quitever4),NA,
                             ifelse(!is.na(ADHbuild$quitever1)&ADHbuild$quitever1==1)|
                                     (!is.na(ADHbuild$quitever2)&ADHbuild$quitever2==1)|
                                     (!is.na(ADHbuild$quitever3)&ADHbuild$quitever3==1)|
                                     (!is.na(ADHbuild$quitever4)&ADHbuild$quitever4==1),1,0))

# ---- ADH: Remove missing tobever, sex, and zygosity -----

ADHbuild <- ADHbuild[-which(is.na(ADHbuild$tobever)),]
table(ADHbuild$zygroup, useNA = 'always') # 86 still missing
ADHbuild <- ADHbuild[-which(is.na(ADHbuild$zygroup)),]
table(ADHbuild$sex, useNA = 'always') # 20 still missing
ADHbuild <- ADHbuild[-which(is.na(ADHbuild$sex)),]

```

```

# ----- ADH: recode conflicting maxcigxday -----
table(ADHbuild$tobever, ADHbuild$maxcigxday, useNA = 'always') # 23 conflicting maxcigxday
ADHbuild$tobever[ADHbuild$tobever == 0 & ADHbuild$maxcigxday > 0] <- 1
ADHbuild$maxcigxdayCCC <- ifelse(ADHbuild$tobever==0, NA, ADHbuild$maxcigxday)
ADHbuild$maxcigxdayCCC <- ifelse(ADHbuild$maxcigxdayCCC == 0, 0,
                                ifelse(ADHbuild$maxcigxdayCCC == 1, 1, 2))
# ----- ADH: recode conflicting quitever -----
table(ADHbuild$tobever, ADHbuild$quitever, useNA = 'always') # 1 conflicting quitever
ADHbuild$tobever[ADHbuild$tobever == 0 & ADHbuild$quitever > 0] <- 1
ADHbuild$quiteverCCC <- ifelse(ADHbuild$tobever==0, NA, ADHbuild$quitever)
# ---- ADH: final dataset -----
ADHbuildfull <- ADHbuild
ADHbuild <- subset(ADHbuildfull, select = c("FAMNO", "IndId", "IndId_int", "StudyNum", "StudyName",
                                           "zygroup", "sex", "ethnicity", "w1age", "w2age", "w3age", "w4age",
                                           "tobever",
                                           "maxcigxdayCCC", "quiteverCCC"))
# ---- ADH: Group by age -----
# First, creat new age variables in each of the waves
# (age11 refers to the age of twin 1 at wave 1, age21 refers to the age of twin 1 at wave 2, , etc.)

ADHbuild$age11 <- ifelse(ADHbuild$IndId_int %in% 1 & !is.na(ADHbuild$w1age), ADHbuild$w1age, NA)
ADHbuild$age21 <- ifelse(ADHbuild$IndId_int %in% 1 & !is.na(ADHbuild$w2age), ADHbuild$w2age, NA)
ADHbuild$age31 <- ifelse(ADHbuild$IndId_int %in% 1 & !is.na(ADHbuild$w3age), ADHbuild$w3age, NA)
ADHbuild$age41 <- ifelse(ADHbuild$IndId_int %in% 1 & !is.na(ADHbuild$w4age), ADHbuild$w4age, NA)

ADHbuild$age12 <- ifelse(ADHbuild$IndId_int %in% 2 & !is.na(ADHbuild$w1age), ADHbuild$w1age, NA)
ADHbuild$age22 <- ifelse(ADHbuild$IndId_int %in% 2 & !is.na(ADHbuild$w2age), ADHbuild$w2age, NA)
ADHbuild$age32 <- ifelse(ADHbuild$IndId_int %in% 2 & !is.na(ADHbuild$w3age), ADHbuild$w3age, NA)
ADHbuild$age42 <- ifelse(ADHbuild$IndId_int %in% 2 & !is.na(ADHbuild$w4age), ADHbuild$w4age, NA)

```

then merged all 4 waves of data into one file and created a new ID variable (AID_1 and AID_2 in the code below)

```
ADHbuild$AID_1 <- ifelse(ADHbuild$IndId_int %in% 1 & !is.na(ADHbuild$IndId_int), 1, NA)
```

```
ADHbuild$AID_2 <- ifelse(ADHbuild$IndId_int %in% 2 & !is.na(ADHbuild$IndId_int), 2, NA)
```

and then based on each participant's age at each wave (i.e. age21 or age11) I created age groups.

If a participant had data for more than one wave, I used the age (and data) from the wave when they were the oldest. So, in the code below I show that I start with age21 and only use age11 if that participant had no data at wave 2.

Participants who had data at more than one wave could appear in the age groups more than once, not within an age group though, but between age groups.

----- ADH: Age group 12-15 -----

```
ADHbuild$agegp12151 <- ifelse(ADHbuild$age41 %in% c(12:15) & !is.na(ADHbuild$AID_1), 1,  
  ifelse(ADHbuild$age31 %in% c(12:15) & !is.na(ADHbuild$AID_1), 1,  
    ifelse(ADHbuild$age21 %in% c(12:15) & !is.na(ADHbuild$AID_1), 1,  
      ifelse(ADHbuild$age11 %in% c(12:15) & !is.na(ADHbuild$AID_1), 1, 0))))
```

```
ADHbuild$agegp12152 <- ifelse(ADHbuild$age42 %in% c(12:15) & !is.na(ADHbuild$AID_2), 1,  
  ifelse(ADHbuild$age32 %in% c(12:15) & !is.na(ADHbuild$AID_2), 1,  
    ifelse(ADHbuild$age22 %in% c(12:15) & !is.na(ADHbuild$AID_2), 1,  
      ifelse(ADHbuild$age12 %in% c(12:15) & !is.na(ADHbuild$AID_2), 1, 0))))
```

----- ADH: Age group 16-19 -----

```
ADHbuild$agegp16191 <- ifelse(ADHbuild$age41 %in% c(16:19) & !is.na(ADHbuild$AID_1), 1,  
  ifelse(ADHbuild$age31 %in% c(16:19) & !is.na(ADHbuild$AID_1), 1,  
    ifelse(ADHbuild$age21 %in% c(16:19) & !is.na(ADHbuild$AID_1), 1,  
      ifelse(ADHbuild$age11 %in% c(16:19) & !is.na(ADHbuild$AID_1), 1, 0))))
```

```
ADHbuild$agegp16192 <- ifelse(ADHbuild$age42 %in% c(16:19) & !is.na(ADHbuild$AID_2), 1,  
  ifelse(ADHbuild$age32 %in% c(16:19) & !is.na(ADHbuild$AID_2), 1,  
    ifelse(ADHbuild$age22 %in% c(16:19) & !is.na(ADHbuild$AID_2), 1,  
      ifelse(ADHbuild$age12 %in% c(16:19) & !is.na(ADHbuild$AID_2), 1, 0))))
```

----- ADH: Age group 22-32 -----

```
ADHbuild$agegp22321 <- ifelse(ADHbuild$age41 %in% c(22:32) & !is.na(ADHbuild$AID_1), 1,  
  ifelse(ADHbuild$age31 %in% c(22:32) & !is.na(ADHbuild$AID_1), 1,
```



```

        ifelse(ADHbuild$age21 %in% c(22:32) & !is.na(ADHbuild$AID_1), 1,
              ifelse(ADHbuild$age11 %in% c(22:32) & !is.na(ADHbuild$AID_1), 1, 0)))
ADHbuild$agegp22322 <- ifelse(ADHbuild$age42 %in% c(22:32) & !is.na(ADHbuild$AID_2), 1,
                          ifelse(ADHbuild$age32 %in% c(22:32) & !is.na(ADHbuild$AID_2), 1,
                                ifelse(ADHbuild$age22 %in% c(22:32) & !is.na(ADHbuild$AID_2), 1,
                                      ifelse(ADHbuild$age12 %in% c(22:32) & !is.na(ADHbuild$AID_2), 1, 0))))

# ---- ADH: Subset age groups ----
ADHG1215 <- ADHbuild[ADHbuild$agegp12151 == 1 | ADHbuild$agegp12152 == 1, ]
ADHG1619 <- ADHbuild[ADHbuild$agegp16191 == 1 | ADHbuild$agegp16192 == 1, ]
ADHG2232 <- ADHbuild[ADHbuild$agegp22321 == 1 | ADHbuild$agegp22322 == 1, ]

write.csv(ADHbuild, "ADHbuild.csv")

write.csv(ADHG1215, "ADHGroup1215.csv")
write.csv(ADHG1619, "ADHGroup1619.csv")
write.csv(ADHG2232, "ADHGroup2232.csv")

# ---- NCT ----
NCT <- read.table(file="NCTquits.csv", header=TRUE, sep=",")
NCTbuild <- NCT

table(NCTbuild$KQ32, useNA = 'always')

# KQ32, if 0 or 1, give a 0 (because 0 = none, 1 = is less than 1 cigs/day)
# if kq32: 2, 3, 4, give a 1 (because, 2 = smoking 1-2/day, 3 = 3-7/day, 4 = 8-12/ day)
# if kq32: 5 or 6, give a 2 (because 5 = 13 to 17, 6 = 18-22)
# 7 or 8, gets a 3 (because 7 = 23-27 cigs/day, 8 = 28-32/day)
# 9 or above, gets a 4 (because 9 = >32/day)
NCTbuild$maxcigxday <- ifelse (NCTbuild$KQ32 == 0 | NCTbuild$KQ32 == 1, 0,
                             ifelse (NCTbuild$KQ32 == 2 | NCTbuild$KQ32 == 3 | NCTbuild$KQ32 == 4, 1,
                                     ifelse (NCTbuild$KQ32 == 5 | NCTbuild$KQ32 == 6, 2,
                                             ifelse (NCTbuild$KQ32 == 7 | NCTbuild$KQ32 == 8, 3,
                                                    ifelse (NCTbuild$KQ32 == 9, 4, NA))))))

# ---- NCT: recode sex for zygoty = 5 ----

```

```

# make male twins as twin 1, and female twins as twin 2
NCTbuild$IndId_int <- ifelse(NCTbuild$zygroup == 5 & !is.na(NCTbuild$sex) & NCTbuild$sex == 1, 1,
                           ifelse (NCTbuild$zygroup == 5 & !is.na(NCTbuild$sex) & NCTbuild$sex == 2, 2,
NCTbuild$IndId))
# ----- NCT: check quitever -----
NCTbuild$quitever <- ifelse(NCTbuild$KQ33==0,0,1)
# ----- NCT: Remove missing tobever, sex, and zygoty -----
table(NCTbuild$tobever, useNA = 'always') #13 missing
NCTbuild <- NCTbuild[-which(is.na(NCTbuild$tobever)),]
table(NCTbuild$zygroup, useNA = 'always') # 113 still missing
NCTbuild <- NCTbuild[-which(is.na(NCTbuild$zygroup)),]
table(NCTbuild$sex, useNA = 'always') # 16 still missing
NCTbuild <- NCTbuild[-which(is.na(NCTbuild$sex)),]
# ----- NCT: recode conflicting maxcigxday -----
table(NCTbuild$tobever, NCTbuild$maxcigxday, useNA = 'always') # 5 conflicting maxcigxday
NCTbuild$tobever[NCTbuild$tobever == 0 & NCTbuild$maxcigxday > 0] <- 1
NCTbuild$maxcigxdayCCC <- ifelse(NCTbuild$tobever==0, NA, NCTbuild$maxcigxday)
NCTbuild$maxcigxdayCCC <- ifelse(NCTbuild$maxcigxdayCCC == 0, 0,
                               ifelse(NCTbuild$maxcigxdayCCC == 1, 1, 2))
# ----- NCT: recode quitever -----
table(NCTbuild$tobever, NCTbuild$quitever, useNA = 'always') # 3 conflicting maxcigxday
NCTbuild$tobever[NCTbuild$tobever == 0 & NCTbuild$quitever > 0] <- 1
NCTbuild$quiteverCCC <- ifelse(NCTbuild$tobever==0, NA, NCTbuild$quitever)
# ----- NCT: final build -----
NCTbuildfull <- NCTbuild
NCTbuild <- subset(NCTbuildfull, select = c("FAMNO", "IndId", "IndId_int", "StudyNum", "StudyName",
                                         "zygroup", "sex", "AGE",
                                         #"ethnicity",
                                         "tobever",
                                         "maxcigxdayCCC", "quiteverCCC"))
# ----- NCT: Group by age -----

```

```

NCTbuild$AID_1 <- ifelse(NCTbuild$IndId_int %in% 1 & !is.na(NCTbuild$IndId_int), 1, NA)
NCTbuild$AID_2 <- ifelse(NCTbuild$IndId_int %in% 2 & !is.na(NCTbuild$IndId_int), 2, NA)
# ---- NCT: Subset age groups ----
NCTG1215 <- NCTbuild[NCTbuild$AGE %in% c(12:15), ]
NCTG1619 <- NCTbuild[NCTbuild$AGE %in% c(16:19), ]

write.csv(NCTbuild, "NCTbuild.csv")
write.csv(NCTG1215, "NCTGroup1215.csv")
write.csv(NCTG1619, "NCTGroup1619.csv")
# ---- Combine QuitBuild ----
#StudyNum == 1: "ABD", StudyNum == 5: "NCT", studynum = 12: "ADH"
# ADH all: ADHbuild
# ADH 1215: ADHG1215
# ADH 1619: ADHG1619
# ADH 2232: ADHG2232
# NCT all: NCTbuild
# NCT 1215: NCTG1215
# NCT 1619: NCTG1619
# ABD all: ABDbuild
# ABD TSA: ABDTSA

ADHbuild$groupID <- 1
ADHG1215$groupID <- 2
ADHG1619$groupID <- 3
ADHG2232$groupID <- 4
NCTbuild$groupID <- 5
NCTG1215$groupID <- 6
NCTG1619$groupID <- 7
ABDbuild$groupID <- 8
ABDTSA$groupID <- 9

```

```

ADHbuild_forQB <- subset(ADHbuild, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

ADHG1215_forQB <- subset(ADHG1215, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

ADHG1619_forQB <- subset(ADHG1619, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

ADHG2232_forQB <- subset(ADHG2232, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

NCTbuild_forQB <- subset(NCTbuild, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

NCTG1215_forQB <- subset(NCTG1215, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

NCTG1619_forQB <- subset(NCTG1619, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

ABDbuild_forQB <- subset(ABDbuild, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

ABDTSA_forQB <- subset(ABDTSA, select = c("FAMNO", "IndId", "IndId_int", "zygroup", "sex",
      "StudyNum", "StudyName", "groupID",
      "tobever", "maxcigxdayCCC", "quiteverCCC"))

QuitBuild <- rbind(ADHbuild_forQB, #1
      ADHG1215_forQB, #2

```

```

ADHG1619_forQB, #3
ADHG2232_forQB, #4
NCTbuild_forQB, #5
NCTG1215_forQB, #6
NCTG1619_forQB, #7
ABDbuild_forQB, #8
ABDTSA_forQB) #9

# ADH
QuitBuild$D1<-ifelse(QuitBuild$groupID==1,1,
  ifelse(QuitBuild$groupID==2,2,
    ifelse(QuitBuild$groupID==3,3,
      ifelse(QuitBuild$groupID==4,4,NA))))

# NCT
QuitBuild$D2<-ifelse(QuitBuild$groupID==5,1,
  ifelse(QuitBuild$groupID==6,2,
    ifelse(QuitBuild$groupID==7,3,NA)))

# ABD
QuitBuild$D3<-ifelse(QuitBuild$groupID==8,1,
  ifelse(QuitBuild$groupID==9, 2, NA))

write.table(QuitBuild, file="QuitBuild_Apr6.csv", row.names=FALSE, col.names=TRUE, sep = ",")

```

Appendix II R script for CCC Model for Add Health age 12-15. (ADH: Add Health)

```
# -----Load Library-----

require(OpenMx)

require(psych)

library('polycor')

mxOption( NULL, "Default optimizer", "NPSOL" )

# -----

# ----- PREPARE DATA -----

data <- read.table("ADHGroup1215.csv", header=TRUE, sep=",")

twindat <- function(dat, famid, twinid, zygosity) {

  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2

  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_ T1", "_ T2"))

  DAT[,paste(twinid, "_ T1", sep="")] <- NULL

  DAT[,paste(twinid, "_ T2", sep="")] <- NULL

  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_ T1", sep="")]), DAT[,paste(zygosity, "_ T2", sep="")], DAT[,paste(zygosity, "_ T1", sep="")])

  DAT[,paste(zygosity, "_ T1", sep="")] <- NULL

  DAT[,paste(zygosity, "_ T2", sep="")] <- NULL

  return(DAT)

}

# three arguments needed for twindat() function

nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")

# Load Data

describe(nic, skew=F)

# Select Variables for Analysis

Vars <- c('tobeaver', 'maxcigxdayCCC', 'quiteverCCC')

nv <- 3 # number of variables

ntv <- nv*2 # number of total variables

selVars <- paste(Vars, c(rep("_ T1", nv), rep("_ T2", nv)), sep="")

# Specify Thresholds for Ordinal Variables
```

```

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1

nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))

thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)

thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

# Heterogeneity CCC ACE Model

```

```

# -----
# Specify Arguments for Causal Path Matrices

svB    <- 0
freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")
labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )
covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )
covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )
covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )
covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )
covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvI <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )
nvZ <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )

```



```

I2    <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1  <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm    <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf    <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm  <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf  <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables

Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds

Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

expCovMZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression= (rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf)))) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af)))+cm%*%t(cf)),
cbind(0.5%x%(rg*(af%*%t(am)))+cf%*%t(cm) , Af+Cf+Ef)), name="expCovDZo" )

```

```

# Data objects for Multiple Groups

dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups

expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")

expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")

expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")

expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")

expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo")

funML <- mxFitFunctionML()

# Combine Groups

pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )

parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)

parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)

modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelADH1215 <- mxModel( "ccc_ADH1215", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ----- RUN GENETIC MODEL -----

# Run CCC and Cholesky Decomposition ACE model

```

```

cccFitADH1215 <- mxTryHard(cccModelADH1215, intervals=F) #minimum
cccFitADH1215_2 <- mxTryHard(cccFitADH1215, intervals=F)
cccFitADH1215_3 <- mxTryHard(cccFitADH1215_2, intervals=F)
cccFitADH1215_4 <- mxTryHard(cccFitADH1215_3, intervals=F)
est_cccFitADH1215 <- summary(cccFitADH1215)
write.csv(est_cccFitADH1215$parameters, "est_cccFitADH1215.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----
cccAceNorgModel_ADH1215 <- mxModel(cccFitADH1215, name="cccAceNorg_ADH1215")
cccAceNorgModel_ADH1215 <- omxSetParameters( cccAceNorgModel_ADH1215, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )
cccAceNorgFit_ADH1215 <- mxTryHard(cccAceNorgModel_ADH1215)
cccAceNorgFit_ADH1215_2 <- mxTryHard(cccAceNorgFit_ADH1215)
cccAceNorgFit_ADH1215_3 <- mxTryHard(cccAceNorgFit_ADH1215_2)
est_cccAceNorgFitADH1215 <- summary(cccAceNorgFit_ADH1215)
write.csv(est_cccAceNorgFitADH1215$parameters, "est_cccAceNorgFitADH1215.csv")
# ----THm=THf----
cccAceNothModel_ADH1215 <- mxModel(cccFitADH1215, name="cccAceNoth_ADH1215")
cccAceNothModel_ADH1215 <- omxSetParameters( cccAceNothModel_ADH1215,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))
cccAceNothModel_ADH1215 <- omxSetParameters( cccAceNothModel_ADH1215,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))
cccAceNothFit_ADH1215 <- mxTryHard(cccAceNothModel_ADH1215)
cccAceNothFit_ADH1215_2 <- mxTryHard(cccAceNothFit_ADH1215)
cccAceNothFit_ADH1215_3 <- mxTryHard(cccAceNothFit_ADH1215_2)
cccAceNothFit_ADH1215_4 <- mxTryHard(cccAceNothFit_ADH1215_3)
cccAceNothFit_ADH1215_5 <- mxTryHard(cccAceNothFit_ADH1215_4)
est_cccAceNothFitADH1215 <- summary(cccAceNothFit_ADH1215_5)
write.csv(est_cccAceNothFitADH1215$parameters, "est_cccAceNothFitADH1215.csv")

```

```

# ---- equal ACE estimate across sex ----

# equate rg, beta, and ACE across sex

cccAceModel_ADH1215 <- mxModel(cccFitADH1215, name="cccAce_ADH1215")

cccAceModel_ADH1215 <- omxSetParameters( cccAceModel_ADH1215, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("af11", "af21", "af31", "af22", "af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("cf11", "cf21", "cf31", "cf22", "cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels=c("ef11", "ef21", "ef31", "ef22", "ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceModel_ADH1215 <- omxSetParameters (cccAceModel_ADH1215, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceFit_ADH1215 <- mxTryHard(cccAceModel_ADH1215)

cccAceFit_ADH1215_2 <- mxTryHard(cccAceFit_ADH1215)

cccAceFit_ADH1215_3 <- mxTryHard(cccAceFit_ADH1215_2)

est_cccAceFitADH1215 <- summary(cccAceFit_ADH1215_3)

write.csv(est_cccAceFitADH1215$parameters, "est_cccAceFitADH1215.csv")

# ---- ACE with No beta ----

cccAceNoBModel_ADH1215 <- mxModel(cccAceFit_ADH1215_3, name="cccAceNoB_ADH1215")

cccAceNoBModel_ADH1215 <- omxSetParameters (cccAceNoBModel_ADH1215, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_ADH1215 <- mxTryHard(cccAceNoBModel_ADH1215)

cccAceNoBFit_ADH1215_2 <- mxTryHard(cccAceNoBFit_ADH1215)

cccAceNoBFit_ADH1215_3 <- mxTryHard(cccAceNoBFit_ADH1215_2)

est_cccAceNoBFitADH1215 <- summary(cccAceNoBFit_ADH1215_3)

write.csv(est_cccAceNoBFitADH1215$parameters, "est_cccAceNoBFitADH1215.csv")

```

```

# ---- ACE drop b_21 ----
cccAceNoB21Model_ADH1215 <- mxModel(cccAceFit_ADH1215_3, name="cccAceNoB21_ADH1215")
cccAceNoB21Model_ADH1215 <- omxSetParameters (cccAceNoB21Model_ADH1215, labels=c("b_21"), free = FALSE,
values = 0)
cccAceNoB21Fit_ADH1215 <- mxTryHard(cccAceNoB21Model_ADH1215)
cccAceNoB21Fit_ADH1215_2 <- mxTryHard(cccAceNoB21Fit_ADH1215)
cccAceNoB21Fit_ADH1215_3 <- mxTryHard(cccAceNoB21Fit_ADH1215_2)
est_cccAceNo21BFitADH1215 <- summary(cccAceNoB21Fit_ADH1215_3)
write.csv(est_cccAceNo21BFitADH1215$parameters, "est_cccAceNo21BFitADH1215.csv")

# ---- ACE drop b_31 ----
cccAceNoB31Model_ADH1215 <- mxModel(cccAceFit_ADH1215_3, name="cccAceNoB31_ADH1215")
cccAceNoB31Model_ADH1215 <- omxSetParameters (cccAceNoB31Model_ADH1215, labels=c("b_31"), free = FALSE,
values = 0)
cccAceNoB31Fit_ADH1215 <- mxTryHard(cccAceNoB31Model_ADH1215)
cccAceNoB31Fit_ADH1215_2 <- mxTryHard(cccAceNoB31Fit_ADH1215)
cccAceNoB31Fit_ADH1215_3 <- mxTryHard(cccAceNoB31Fit_ADH1215_2)
est_cccAceNo31BFitADH1215 <- summary(cccAceNoB31Fit_ADH1215_3)
write.csv(est_cccAceNo31BFitADH1215$parameters, "est_cccAceNo31BFitADH1215.csv")

# ---- ACE drop b31, drop C ----
cccAeNoB31Model_ADH1215 <- mxModel(cccAceFit_ADH1215_3, name="cccAeNoB31_ADH1215")
cccAeNoB31Model_ADH1215 <- omxSetParameters (cccAeNoB31Model_ADH1215, labels=c("b_31"), free = FALSE, values
= 0)
cccAeNoB31Model_ADH1215 <- omxSetParameters (cccAeNoB31Model_ADH1215, labels=c("c11", "c22", "c32", "c33"), free
= FALSE, values = 0)
cccAeNoB31Fit_ADH1215 <- mxTryHard(cccAeNoB31Model_ADH1215)
cccAeNoB31Fit_ADH1215_2 <- mxTryHard(cccAeNoB31Fit_ADH1215)
cccAeNoB31Fit_ADH1215_3 <- mxTryHard(cccAeNoB31Fit_ADH1215_2)
est_cccAeNo31BFitADH1215 <- summary(cccAeNoB31Fit_ADH1215_3)
write.csv(est_cccAeNo31BFitADH1215$parameters, "est_cccAeNo31BFitADH1215.csv")

# ---- ACE drop beta 31, drop A ----
cccCeNoB31Model_ADH1215 <- mxModel(cccAceFit_ADH1215_3, name="cccCeNoB31_ADH1215")
cccCeNoB31Model_ADH1215 <- omxSetParameters (cccCeNoB31Model_ADH1215, labels=c("b_31"), free = FALSE, values
= 0)
cccCeNoB31Model_ADH1215 <- omxSetParameters (cccCeNoB31Model_ADH1215, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)

```

```

cccCeNoB31Fit_ADH1215 <- mxTryHard(cccCeNoB31Model_ADH1215)
cccCeNoB31Fit_ADH1215_2 <- mxTryHard(cccCeNoB31Fit_ADH1215)
cccCeNoB31Fit_ADH1215_3 <- mxTryHard(cccCeNoB31Fit_ADH1215_2)
est_cccCeNo31BFitADH1215 <- summary(cccCeNoB31Fit_ADH1215_3)
write.csv(est_cccCeNo31BFitADH1215$parameters, "est_cccCeNo31BFitADH1215.csv")
# ---- All submodels ----
fitstatsubmodels1_ADH1215 <- mxCompare(cccFitADH1215, nested <- list(cccAceNorgFit_ADH1215,
                                                                    cccAceNothFit_ADH1215_5,
                                                                    cccAceFit_ADH1215_3))
fitstatsubmodels2_ADH1215 <- mxCompare(cccAceFit_ADH1215_3, nested <- list(cccAceNoBFit_ADH1215_3,
                                                                    cccAceNoB21Fit_ADH1215_3,
                                                                    cccAceNoB31Fit_ADH1215_3,
                                                                    cccAeNoB31Fit_ADH1215_3,
                                                                    cccCeNoB31Fit_ADH1215_3))
fitstatsubmodels_ADH1215 <- rbind(fitstatsubmodels1_ADH1215, fitstatsubmodels2_ADH1215)
write.csv(fitstatsubmodels_ADH1215, "submodelsADH1215.csv")
# base model
write.csv(summary(cccFitADH1215)$parameters, "estimates_basemodel_ADH_1215.csv")
# ---- Create Confidence Interval Objects: best fitting model ----
# run the best fitting with 95% CIs
bestci <- mxCI(c("a11", "a22", "a32", "a33",
                "e11", "e22", "e32", "e33",
                "b_21"))
bestmodel_ADH_1215 <- mxModel(cccAeNoB31Fit_ADH1215_3, bestci)
fitbest_ADH_1215 <- mxTryHard(bestmodel_ADH_1215, intervals = F)
fitbest_ADH_1215_2 <- mxTryHard(fitbest_ADH_1215, intervals = F)
fitbest_ADH_1215_3 <- mxTryHard(fitbest_ADH_1215_2, intervals = F)
fitbest_ADH_1215_int <- mxRun(fitbest_ADH_1215_3, intervals = T)
sumbest_ADH_1215 <- summary(fitbest_ADH_1215_int)
sumbest_ADH_1215$parameters
sumbest_ADH_1215$CI

```

```

sumbest_ADH_1215$CI$CIint <- paste("(",round(sumbest_ADH_1215$CI$lbound,2), "-",
round(sumbest_ADH_1215$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_ADH_1215$CI, "CI_bestfitting_ADH_1215.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                        "c11", "c22", "c32", "c33",
                        "e11", "e22", "e32", "e33",
                        "b_21", "b_31"))

ACEmodel_ADH_1215 <- mxModel(cccAceFit_ADH1215_3, ACEciACEnorg)

fitACE_ADH_1215 <- mxTryHard(ACEmodel_ADH_1215, intervals = F)

fitACE_ADH_1215_2 <- mxTryHard(fitACE_ADH_1215, intervals = F)

fitACE_ADH_1215_3 <- mxTryHard(fitACE_ADH_1215_2, intervals = F)

fitACE_ADH_1215_int <- mxRun(fitACE_ADH_1215_3, intervals = T)

sumACE_ADH_1215 <- summary(fitACE_ADH_1215_int)

sumACE_ADH_1215$parameters

sumACE_ADH_1215$CI

sumACE_ADH_1215$CI$CIint <- paste("(",round(sumACE_ADH_1215$CI$lbound,2), "-",
round(sumACE_ADH_1215$CI$subound,2),")", sep="")

sumACE_ADH_12151 <- sumACE_ADH_1215$CI

sumACE_ADH_12151$datagroup <- "ADH_1215"

write.csv(sumACE_ADH_12151, "sumACE_ADH_1215.csv")

```

Appendix III R script for CCC Model for Add Health age 16-19. (ADH: Add Health)

```
# -----Load Library-----
require(OpenMx)
require(psych)
source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )
# -----
# ----- PREPARE DATA -----
data <- read.table("ADHGroup1619.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobever', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```



```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1
nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))
thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)
thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )

covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )

covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )

covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )

covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvl   <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvl" )

```

```

nvZ    <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2     <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1   <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm     <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf     <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm   <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf   <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables
Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds
Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )
Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )
expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix
expCovMZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression=( rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf)))) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af))+cm%*%t(cf)),

```

```

cbind(0.5*x%(rg*(af%*%(am)))+cf%*%(cm) , Af+Cf+Ef), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo")

funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelADH1619 <- mxModel( "ccc_ADH1619", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ----- RUN GENETIC MODEL -----

# Run CCC and Cholesky Decomposition ACE model

```

```

cccFitADH1619 <- mxTryHard(cccModelADH1619, intervals=F)
cccFitADH1619_2 <- mxTryHard(cccFitADH1619, intervals=F)
cccFitADH1619_3 <- mxTryHard(cccFitADH1619_2, intervals=F)
cccFitADH1619_4 <- mxTryHard(cccFitADH1619_3, intervals=F)
est_cccFitADH1619 <- summary(cccFitADH1619_4)
write.csv(est_cccFitADH1619$parameters, "est_cccFitADH1619.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----

cccAceNorgModel_ADH1619 <- mxModel(cccFitADH1619_4, name="cccAceNorg_ADH1619")
cccAceNorgModel_ADH1619 <- omxSetParameters( cccAceNorgModel_ADH1619, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )
cccAceNorgFit_ADH1619 <- mxTryHard(cccAceNorgModel_ADH1619)
cccAceNorgFit_ADH1619_2 <- mxTryHard(cccAceNorgFit_ADH1619)
cccAceNorgFit_ADH1619_3 <- mxTryHard(cccAceNorgFit_ADH1619_2)
cccAceNorgFit_ADH1619_4 <- mxTryHard(cccAceNorgFit_ADH1619_3)
cccAceNorgFit_ADH1619_5 <- mxTryHard(cccAceNorgFit_ADH1619_4)
est_cccAceNorgFitADH1619 <- summary(cccAceNorgFit_ADH1619_3)
write.csv(est_cccAceNorgFitADH1619$parameters, "est_cccAceNorgFitADH1619.csv")

# ---- equal ACE estimate across sex ----
# equate rg, beta, and ACE across sex

cccAceModel_ADH1619 <- mxModel(cccFitADH1619_4, name="cccAce_ADH1619")
cccAceModel_ADH1619 <- omxSetParameters( cccAceModel_ADH1619, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("am11","am21","am31", "am22","am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("cm11","cm21","cm31", "cm22","cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("em11","em21","em31", "em22","em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("af11","af21","af31", "af22","af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("cf11","cf21","cf31", "cf22","cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )
cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels=c("ef11","ef21","ef31", "ef22","ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )

```

```

cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceModel_ADH1619 <- omxSetParameters (cccAceModel_ADH1619, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceFit_ADH1619 <- mxTryHard(cccAceModel_ADH1619)

cccAceFit_ADH1619_2 <- mxTryHard(cccAceFit_ADH1619)

cccAceFit_ADH1619_3 <- mxTryHard(cccAceFit_ADH1619_2)

cccAceFit_ADH1619_4 <- mxTryHard(cccAceFit_ADH1619_3)

#cccAceFit_ADH1619_5 <- mxTryHard(cccAceFit_ADH1619_4)

est_cccAceFitADH1619 <- summary(cccAceFit_ADH1619_4)

write.csv(est_cccAceFitADH1619$parameters, "est_cccAceFitADH1619.csv")

# ----THm=THf-----

cccAceNothModel_ADH1619 <- mxModel(cccFitADH1619_4, name="cccAceNoth_ADH1619")

cccAceNothModel_ADH1619 <- omxSetParameters( cccAceNothModel_ADH1619,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothModel_ADH1619 <- omxSetParameters( cccAceNothModel_ADH1619,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothFit_ADH1619 <- mxTryHard(cccAceNothModel_ADH1619)

cccAceNothFit_ADH1619_2 <- mxTryHard(cccAceNothFit_ADH1619)

cccAceNothFit_ADH1619_3 <- mxTryHard(cccAceNothFit_ADH1619_2)

cccAceNothFit_ADH1619_4 <- mxTryHard(cccAceNothFit_ADH1619_3)

est_cccAceNothFitADH1619 <- summary(cccAceNothFit_ADH1619_4)

write.csv(est_cccAceNothFitADH1619$parameters, "est_cccAceNothFitADH1619.csv")

# ---- ACE with No beta -----

cccAceNoBModel_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccAceNoB_ADH1619")

cccAceNoBModel_ADH1619 <- omxSetParameters (cccAceNoBModel_ADH1619, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_ADH1619 <- mxTryHard(cccAceNoBModel_ADH1619)

cccAceNoBFit_ADH1619_2 <- mxTryHard(cccAceNoBFit_ADH1619)

cccAceNoBFit_ADH1619_3 <- mxTryHard(cccAceNoBFit_ADH1619_2)

est_cccAceNoBFitADH1619 <- summary(cccAceNoBFit_ADH1619_3)

```

```

write.csv(est_cccAceNoBFitADH1619$parameters, "est_cccAceNoBFitADH1619.csv")

# ---- ACE drop b_21 ----

cccAceNoB21Model_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccAceNoB21_ADH1619")

cccAceNoB21Model_ADH1619 <- omxSetParameters (cccAceNoB21Model_ADH1619, labels=c("b_21"), free = FALSE,
values = 0)

cccAceNoB21Fit_ADH1619 <- mxTryHard(cccAceNoB21Model_ADH1619)

cccAceNoB21Fit_ADH1619_2 <- mxTryHard(cccAceNoB21Fit_ADH1619)

cccAceNoB21Fit_ADH1619_3 <- mxTryHard(cccAceNoB21Fit_ADH1619_2)

est_cccAceNo21BFitADH1619 <- summary(cccAceNoB21Fit_ADH1619_3)

write.csv(est_cccAceNo21BFitADH1619$parameters, "est_cccAceNo21BFitADH1619.csv")

# ---- ACE drop b_31 ----

cccAceNoB31Model_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccAceNoB31_ADH1619")

cccAceNoB31Model_ADH1619 <- omxSetParameters (cccAceNoB31Model_ADH1619, labels=c("b_31"), free = FALSE,
values = 0)

cccAceNoB31Fit_ADH1619 <- mxTryHard(cccAceNoB31Model_ADH1619)

cccAceNoB31Fit_ADH1619_2 <- mxTryHard(cccAceNoB31Fit_ADH1619)

cccAceNoB31Fit_ADH1619_3 <- mxTryHard(cccAceNoB31Fit_ADH1619_2)

est_cccAceNo31BFitADH1619 <- summary(cccAceNoB31Fit_ADH1619_3)

write.csv(est_cccAceNo31BFitADH1619$parameters, "est_cccAceNo31BFitADH1619.csv")

# ---- ACE drop b31, drop C ----

cccAeNoB31Model_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccAeNoB31_ADH1619")

cccAeNoB31Model_ADH1619 <- omxSetParameters (cccAeNoB31Model_ADH1619, labels=c("b_31"), free = FALSE, values
= 0)

cccAeNoB31Model_ADH1619 <- omxSetParameters (cccAeNoB31Model_ADH1619, labels=c("c11", "c22", "c32", "c33"), free
= FALSE, values = 0)

cccAeNoB31Fit_ADH1619 <- mxTryHard(cccAeNoB31Model_ADH1619)

cccAeNoB31Fit_ADH1619_2 <- mxTryHard(cccAeNoB31Fit_ADH1619)

cccAeNoB31Fit_ADH1619_3 <- mxTryHard(cccAeNoB31Fit_ADH1619_2)

cccAeNoB31Fit_ADH1619_4 <- mxTryHard(cccAeNoB31Fit_ADH1619_3)

est_cccAeNo31BFitADH1619 <- summary(cccAeNoB31Fit_ADH1619)

write.csv(est_cccAeNo31BFitADH1619$parameters, "est_cccAeNo31BFitADH1619.csv")

# ---- ACE drop beta 31, drop A ----

cccCeNoB31Model_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccCeNoB31_ADH1619")

```

```

cccCeNoB31Model_ADH1619 <- omxSetParameters (cccCeNoB31Model_ADH1619, labels=c("b_31"), free = FALSE, values
= 0)

cccCeNoB31Model_ADH1619 <- omxSetParameters (cccCeNoB31Model_ADH1619, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)

cccCeNoB31Fit_ADH1619 <- mxTryHard(cccCeNoB31Model_ADH1619)

cccCeNoB31Fit_ADH1619_2 <- mxTryHard(cccCeNoB31Fit_ADH1619)

cccCeNoB31Fit_ADH1619_3 <- mxTryHard(cccCeNoB31Fit_ADH1619_2)

est_cccCeNo31BFitADH1619 <- summary(cccCeNoB31Fit_ADH1619_3)

write.csv(est_cccCeNo31BFitADH1619$parameters, "est_cccCeNo31BFitADH1619.csv")

# ---- ACE drop beta 31, drop A33 ----

cccACeNoB31A33Model_ADH1619 <- mxModel(cccAceFit_ADH1619_4, name="cccACeNoB31A33_ADH1619")

cccACeNoB31A33Model_ADH1619 <- omxSetParameters (cccACeNoB31A33Model_ADH1619, labels=c("b_31"), free =
FALSE, values = 0)

cccACeNoB31A33Model_ADH1619 <- omxSetParameters (cccACeNoB31A33Model_ADH1619, labels=c("c11", "c22", "c32",
"c33"), free = FALSE, values = 0)

cccACeNoB31A33Model_ADH1619 <- omxSetParameters (cccACeNoB31A33Model_ADH1619, labels=c("a33"), free =
FALSE, values = 0)

cccACeNoB31A33Fit_ADH1619 <- mxTryHard(cccACeNoB31A33Model_ADH1619)

cccACeNoB31A33Fit_ADH1619_2 <- mxTryHard(cccACeNoB31A33Fit_ADH1619)

est_cccACeNoB31A33Fit_ADH1619_2 <- summary(cccACeNoB31A33Fit_ADH1619_2)

write.csv(est_cccACeNoB31A33Fit_ADH1619_2$parameters, "est_cccACeNoB31A33Fit_ADH1619_2.csv")

# ---- All submodels ----

fitstatsubmodels1_ADH1619 <- mxCompare(cccFitADH1619_4, nested <- list(cccAceNorgFit_ADH1619_3,
                                cccAceNothFit_ADH1619_4,
                                cccAceFit_ADH1619_4))

fitstatsubmodels2_ADH1619 <- mxCompare(cccAceFit_ADH1619_4, nested <- list(cccAceNoBFit_ADH1619_3,
                                cccAceNoB21Fit_ADH1619_3,
                                cccAceNoB31Fit_ADH1619_3,
                                cccAeNoB31Fit_ADH1619,
                                cccCeNoB31Fit_ADH1619_3,
                                cccACeNoB31A33Fit_ADH1619_2))

fitstatsubmodels_ADH1619 <- rbind(fitstatsubmodels1_ADH1619, fitstatsubmodels2_ADH1619)

write.csv(fitstatsubmodels_ADH1619, "submodelsADH1619.csv")

```



```

# base model

write.csv(summary(cccFitADH1619_4)$parameters, "estimates_basemodel_ADH_1619.csv")

# ----- Create Confidence Interval Objects: best fitting model -----

# run the best fitting with 95% CIs

bestci  <- mxCI( c("a11", "a22", "a32", "a33",
                  "e11", "e22", "e32", "e33",
                  "b_21"))

bestmodel_ADH_1619 <- mxModel(cccAeNoB31Fit_ADH1619, bestci)

fitbest_ADH_1619 <- mxTryHard(bestmodel_ADH_1619, intervals = F)

fitbest_ADH_1619_2 <- mxTryHard(fitbest_ADH_1619, intervals = F)

fitbest_ADH_1619_3 <- mxTryHard(fitbest_ADH_1619_2, intervals = F)

fitbest_ADH_1619_int <- mxRun(fitbest_ADH_1619_3, intervals = T)

sumbest_ADH_1619 <- summary(fitbest_ADH_1619_int)

sumbest_ADH_1619$parameters

sumbest_ADH_1619$CI

sumbest_ADH_1619$CI$CIint <- paste("(",round(sumbest_ADH_1619$CI$lbound,2), "-",
round(sumbest_ADH_1619$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_ADH_1619$CI, "CI_bestfitting_ADH_1619.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg  <- mxCI( c("a11", "a22", "a32", "a33",
                        "c11", "c22", "c32", "c33",
                        "e11", "e22", "e32", "e33",
                        "b_21", "b_31"))

ACEmodel_ADH_1619 <- mxModel(cccAceFit_ADH1619_4, ACEciACEnorg)

fitACE_ADH_1619 <- mxTryHard(ACEmodel_ADH_1619, intervals = F)

fitACE_ADH_1619_2 <- mxTryHard(fitACE_ADH_1619, intervals = F)

fitACE_ADH_1619_3 <- mxTryHard(fitACE_ADH_1619_2, intervals = F)

fitACE_ADH_1619_int <- mxRun(fitACE_ADH_1619_3, intervals = T)

sumACE_ADH_1619 <- summary(fitACE_ADH_1619_int)

sumACE_ADH_1619$parameters

sumACE_ADH_1619$CI

```

```
sumACE_ADH_1619$CI$CIint <- paste("(",round(sumACE_ADH_1619$CI$lbound,2), "-",  
round(sumACE_ADH_1619$CI$subound,2),")", sep="")  
  
sumACE_ADH_16191 <- sumACE_ADH_1619$CI  
  
sumACE_ADH_16191$datagroup <- "ADH_1619"  
  
write.csv(sumACE_ADH_16191, "sumACE_ADH_1619.csv")
```

Appendix IV R script for CCC Model for Add Health age 22-32. (ADH: Add Health)

```
# -----Load Library-----
require(OpenMx)
require(psych)
source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )
# -----
# ----- PREPARE DATA -----
data <- read.table("ADHGroup2232.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobever', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```

```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1
nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))
thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)
thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )

covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )

covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )

covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )

covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvI    <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )

```

```

nvZ    <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2     <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1   <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm     <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf     <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm   <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf   <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables

Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds

Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

expCovMZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression=( rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf))) ) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af))+cm%*%t(cf)),

```

```

cbind(0.5%x%(rg*(af%*%(am)))+cf%*%(cm) , Af+Cf+Ef), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo")
funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )
multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))
cccModelADH2232 <- mxModel( "ccc_ADH2232", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ----- RUN GENETIC MODEL -----

```

```

# Run CCC and Cholesky Decomposition ACE model
cccFitADH2232 <- mxTryHard(cccModelADH2232, intervals=F)
cccFitADH2232_2 <- mxTryHard(cccFitADH2232, intervals=F)
cccFitADH2232_3 <- mxTryHard(cccFitADH2232_2, intervals=F)
cccFitADH2232_4 <- mxTryHard(cccFitADH2232_3, intervals=F)
cccFitADH2232_5 <- mxTryHard(cccFitADH2232_4, intervals=F)
est_cccFitADH2232 <- summary(cccFitADH2232_2)
write.csv(est_cccFitADH2232$parameters, "est_cccFitADH2232.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----

cccAceNorgModel_ADH2232 <- mxModel(cccFitADH2232_2, name="cccAceNorg_ADH2232")
cccAceNorgModel_ADH2232 <- omxSetParameters( cccAceNorgModel_ADH2232, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )
cccAceNorgFit_ADH2232 <- mxTryHard(cccAceNorgModel_ADH2232)
cccAceNorgFit_ADH2232_2 <- mxTryHard(cccAceNorgFit_ADH2232)
cccAceNorgFit_ADH2232_3 <- mxTryHard(cccAceNorgFit_ADH2232_2)
cccAceNorgFit_ADH2232_4 <- mxTryHard(cccAceNorgFit_ADH2232_3)
#cccAceNorgFit_ADH2232_5 <- mxTryHard(cccAceNorgFit_ADH2232_4)
#cccAceNorgFit_ADH2232_6 <- mxTryHard(cccAceNorgFit_ADH2232_5)
est_cccAceNorgFitADH2232 <- summary(cccAceNorgFit_ADH2232)
write.csv(est_cccAceNorgFitADH2232$parameters, "est_cccAceNorgFitADH2232.csv")

# ----THm=THf-----

cccAceNothModel_ADH2232 <- mxModel(cccFitADH2232_2, name="cccAceNoth_ADH2232")
cccAceNothModel_ADH2232 <- omxSetParameters( cccAceNothModel_ADH2232,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))
cccAceNothModel_ADH2232 <- omxSetParameters( cccAceNothModel_ADH2232,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothFit_ADH2232 <- mxTryHard(cccAceNothModel_ADH2232)
cccAceNothFit_ADH2232_2 <- mxTryHard(cccAceNothFit_ADH2232)

```



```

cccAceNothFit_ADH2232_3 <- mxTryHard(cccAceNothFit_ADH2232_2)
cccAceNothFit_ADH2232_4 <- mxTryHard(cccAceNothFit_ADH2232_3)
est_cccAceNothFitADH2232 <- summary(cccAceNothFit_ADH2232_2)
write.csv(est_cccAceNothFitADH2232$parameters, "est_cccAceNothFitADH2232.csv")

# ---- equal ACE estimate across sex ----

# equate rg, beta, and ACE across sex

cccAceModel_ADH2232 <- mxModel(cccFitADH2232_2, name="cccAce_ADH2232")

cccAceModel_ADH2232 <- omxSetParameters( cccAceModel_ADH2232, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("af11", "af21", "af31", "af22", "af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("cf11", "cf21", "cf31", "cf22", "cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels=c("ef11", "ef21", "ef31", "ef22", "ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceModel_ADH2232 <- omxSetParameters (cccAceModel_ADH2232, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceFit_ADH2232 <- mxTryHard(cccAceModel_ADH2232)

cccAceFit_ADH2232_2 <- mxTryHard(cccAceFit_ADH2232)

cccAceFit_ADH2232_3 <- mxTryHard(cccAceFit_ADH2232_2)

cccAceFit_ADH2232_4 <- mxTryHard(cccAceFit_ADH2232_3)

cccAceFit_ADH2232_5 <- mxTryHard(cccAceFit_ADH2232_4)

est_cccAceFitADH2232 <- summary(cccAceFit_ADH2232_5)

write.csv(est_cccAceFitADH2232$parameters, "est_cccAceFitADH2232.csv")

# ---- ACE with No beta ----

cccAceNoBModel_ADH2232 <- mxModel(cccAceFit_ADH2232_5, name="cccAceNoB_ADH2232")

cccAceNoBModel_ADH2232 <- omxSetParameters (cccAceNoBModel_ADH2232, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

```

```

cccAceNoBFit_ADH2232 <- mxTryHard(cccAceNoBModel_ADH2232)
cccAceNoBFit_ADH2232_2 <- mxTryHard(cccAceNoBFit_ADH2232)
cccAceNoBFit_ADH2232_3 <- mxTryHard(cccAceNoBFit_ADH2232_2)
est_cccAceNoBFitADH2232 <- summary(cccAceNoBFit_ADH2232_3)
write.csv(est_cccAceNoBFitADH2232$parameters, "est_cccAceNoBFitADH2232.csv")
# ---- ACE drop b_21 ----
cccAceNoB21Model_ADH2232 <- mxModel(cccAceFit_ADH2232_5, name="cccAceNoB21_ADH2232")
cccAceNoB21Model_ADH2232 <- omxSetParameters (cccAceNoB21Model_ADH2232, labels=c("b_21"), free = FALSE,
values = 0)
cccAceNoB21Fit_ADH2232 <- mxTryHard(cccAceNoB21Model_ADH2232)
cccAceNoB21Fit_ADH2232_2 <- mxTryHard(cccAceNoB21Fit_ADH2232)
cccAceNoB21Fit_ADH2232_3 <- mxTryHard(cccAceNoB21Fit_ADH2232_2)
est_cccAceNo21BFitADH2232 <- summary(cccAceNoB21Fit_ADH2232_3)
write.csv(est_cccAceNo21BFitADH2232$parameters, "est_cccAceNo21BFitADH2232.csv")
# ---- ACE drop b_31 ----
cccAceNoB31Model_ADH2232 <- mxModel(cccAceFit_ADH2232_5, name="cccAceNoB31_ADH2232")
cccAceNoB31Model_ADH2232 <- omxSetParameters (cccAceNoB31Model_ADH2232, labels=c("b_31"), free = FALSE,
values = 0)
cccAceNoB31Fit_ADH2232 <- mxTryHard(cccAceNoB31Model_ADH2232)
cccAceNoB31Fit_ADH2232_2 <- mxTryHard(cccAceNoB31Fit_ADH2232)
cccAceNoB31Fit_ADH2232_3 <- mxTryHard(cccAceNoB31Fit_ADH2232_2)
est_cccAceNo31BFitADH2232 <- summary(cccAceNoB31Fit_ADH2232_3)
write.csv(est_cccAceNo31BFitADH2232$parameters, "est_cccAceNo31BFitADH2232.csv")
# ---- ACE drop b31, drop C ----
cccAeNoB31Model_ADH2232 <- mxModel(cccAceFit_ADH2232_5, name="cccAeNoB31_ADH2232")
cccAeNoB31Model_ADH2232 <- omxSetParameters (cccAeNoB31Model_ADH2232, labels=c("b_31"), free = FALSE, values
= 0)
cccAeNoB31Model_ADH2232 <- omxSetParameters (cccAeNoB31Model_ADH2232, labels=c("c11", "c22","c32", "c33"), free
= FALSE, values = 0)
cccAeNoB31Fit_ADH2232 <- mxTryHard(cccAeNoB31Model_ADH2232)
cccAeNoB31Fit_ADH2232_2 <- mxTryHard(cccAeNoB31Fit_ADH2232)
cccAeNoB31Fit_ADH2232_3 <- mxTryHard(cccAeNoB31Fit_ADH2232_2)
est_cccAeNo31BFitADH2232 <- summary(cccAeNoB31Fit_ADH2232_3)
write.csv(est_cccAeNo31BFitADH2232$parameters, "est_cccAeNo31BFitADH2232.csv")

```

```

# ---- ACE drop beta 31, drop A ----
cccCeNoB31Model_ADH2232 <- mxModel(cccAceFit_ADH2232_5, name="cccCeNoB31_ADH2232")

cccCeNoB31Model_ADH2232 <- omxSetParameters (cccCeNoB31Model_ADH2232, labels=c("b_31"), free = FALSE, values
= 0)

cccCeNoB31Model_ADH2232 <- omxSetParameters (cccCeNoB31Model_ADH2232, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)

cccCeNoB31Fit_ADH2232 <- mxTryHard(cccCeNoB31Model_ADH2232)

cccCeNoB31Fit_ADH2232_2 <- mxTryHard(cccCeNoB31Fit_ADH2232)

cccCeNoB31Fit_ADH2232_3 <- mxTryHard(cccCeNoB31Fit_ADH2232_2)

est_cccCeNo31BFitADH2232 <- summary(cccCeNoB31Fit_ADH2232_3)

write.csv(est_cccCeNo31BFitADH2232$parameters, "est_cccCeNo31BFitADH2232.csv")

# ---- ACE drop beta 31, drop A33 ----

cccACeNoB31A33Model_ADH2232 <- mxModel(cccAceFit_ADH2232_4, name="cccACeNoB31A33_ADH2232")

cccACeNoB31A33Model_ADH2232 <- omxSetParameters (cccACeNoB31A33Model_ADH2232, labels=c("b_31"), free =
FALSE, values = 0)

cccACeNoB31A33Model_ADH2232 <- omxSetParameters (cccACeNoB31A33Model_ADH2232, labels=c("c11", "c22", "c32",
"c33"), free = FALSE, values = 0)

cccACeNoB31A33Model_ADH2232 <- omxSetParameters (cccACeNoB31A33Model_ADH2232, labels=c("a33"), free =
FALSE, values = 0)

cccACeNoB31A33Fit_ADH2232 <- mxTryHard(cccACeNoB31A33Model_ADH2232)

cccACeNoB31A33Fit_ADH2232_2 <- mxTryHard(cccACeNoB31A33Fit_ADH2232)

est_cccACeNoB31A33Fit_ADH2232_2 <- summary(cccACeNoB31A33Fit_ADH2232_2)

write.csv(est_cccACeNoB31A33Fit_ADH2232_2$parameters, "est_cccACeNoB31A33Fit_ADH2232_2.csv")

# ---- All submodels ----

fitstatsubmodels1_ADH2232 <- mxCompare(cccFitADH2232_2, nested <- list(cccAceNorgFit_ADH2232,
                                cccAceNothFit_ADH2232_2,
                                cccAceFit_ADH2232_5))

fitstatsubmodels2_ADH2232 <- mxCompare(cccAceFit_ADH2232_5, nested <- list(cccAceNoBFit_ADH2232_3,
                                cccAceNoB21Fit_ADH2232_3,
                                cccAceNoB31Fit_ADH2232_3,
                                cccAeNoB31Fit_ADH2232_3,
                                cccCeNoB31Fit_ADH2232_3))

fitstatsubmodels_ADH2232 <- rbind(fitstatsubmodels1_ADH2232, fitstatsubmodels2_ADH2232)

```

```

write.csv(fitstatsubmodels_ADH2232, "submodelsADH2232.csv")

# base model

write.csv(summary(cccFitADH2232_2)$parameters, "estimates_basemodel_ADH_2232.csv")

# ----- Create Confidence Interval Objects: best fitting model -----

# run the best fitting with 95% CIs

bestci <- mxCI( c("a11", "a22", "a32", "a33",
                "e11", "e22", "e32", "e33",
                "b_21"))

bestmodel_ADH_2232 <- mxModel(cccAeNoB31Fit_ADH2232_3, bestci)

fitbest_ADH_2232 <- mxTryHard(bestmodel_ADH_2232, intervals = F)
fitbest_ADH_2232_2 <- mxTryHard(fitbest_ADH_2232, intervals = F)
fitbest_ADH_2232_3 <- mxTryHard(fitbest_ADH_2232_2, intervals = F)
fitbest_ADH_2232_int <- mxRun(fitbest_ADH_2232_3, intervals = T)
sumbest_ADH_2232 <- summary(fitbest_ADH_2232_int)

sumbest_ADH_2232$parameters

sumbest_ADH_2232$CI

sumbest_ADH_2232$CI$CIint <- paste("(",round(sumbest_ADH_2232$CI$lbound,2), "-",
round(sumbest_ADH_2232$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_ADH_2232$CI, "CI_bestfitting_ADH_2232.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                    "c11", "c22", "c32", "c33",
                    "e11", "e22", "e32", "e33",
                    "b_21", "b_31"))

ACEmodel_ADH_2232 <- mxModel(cccAceFit_ADH2232_5, ACEciACEnorg)

fitACE_ADH_2232 <- mxTryHard(ACEmodel_ADH_2232, intervals = F)
fitACE_ADH_2232_2 <- mxTryHard(fitACE_ADH_2232, intervals = F)
fitACE_ADH_2232_3 <- mxTryHard(fitACE_ADH_2232_2, intervals = F)
fitACE_ADH_2232_int <- mxRun(fitACE_ADH_2232_3, intervals = T)
sumACE_ADH_2232 <- summary(fitACE_ADH_2232_int)

sumACE_ADH_2232$parameters

```

```
sumACE_ADH_2232$CI
sumACE_ADH_2232$CI$CIint <- paste("(",round(sumACE_ADH_2232$CI$lbound,2), "-",
round(sumACE_ADH_2232$CI$subound,2),")", sep="")
sumACE_ADH_22321 <- sumACE_ADH_2232$CI
sumACE_ADH_22321$datagroup <- "ADH_2232"
write.csv(sumACE_ADH_22321, "sumACE_ADH_2232.csv")
```

Appendix V R script for CCC Model for Add Health all participants. (ADH: Add Health)

```
# -----Load Library-----
require(OpenMx)
require(psych)
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )

# -----
# ----- PREPARE DATA -----

data <- read.table("ADHbuild.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}

# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zyggroup")

# Load Data
describe(nic, skew=F)

# Select Variables for Analysis
Vars <- c('tobeveer', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```

```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1 <- 1 ; fcat1 <- 0 ; lcat1 <- fcat1+nth1 ; ncat1 <- nth1+1
nth2 <- 2 ; fcat2 <- 0 ; lcat2 <- fcat2+nth2 ; ncat2 <- nth2+1
nth3 <- 1 ; fcat3 <- 0 ; lcat3 <- fcat3+nth3 ; ncat3 <- nth3+1
nth <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1 <- 0 ; ith1 <- 0.1 ;
lth2 <- 0 ; ith2 <- 0.1 ;
lth3 <- 0 ; ith3 <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))
thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)
thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]
describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )

covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )

covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )

covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )

covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvl <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvl" )

```



```

nvZ    <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2     <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1   <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm     <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf     <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm   <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf   <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables
Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds
Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )
Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )
expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix
expCovMZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression=( rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf))) ) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af))+cm%*%t(cf)),

```

```

cbind(0.5*x%(rg*(af%*%(am)))+cf%*%(cm) , Af+Cf+Ef), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo")
funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )
multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))
cccModelADH_all <- mxModel( "ccc_ADH_all", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ----- RUN GENETIC MODEL -----

```

```

# Run CCC and Cholesky Decomposition ACE model

cccFitADH_all <- mxTryHard(cccModelADH_all, intervals=F)
cccFitADH_all_2 <- mxTryHard(cccFitADH_all, intervals=F)
cccFitADH_all_3 <- mxTryHard(cccFitADH_all_2, intervals=F)
est_cccFitADH_all <- summary(cccFitADH_all_3)
write.csv(est_cccFitADH_all$parameters, "est_cccFitADH_all.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----

cccAceNorgModel_ADH_all <- mxModel(cccFitADH_all_3, name="cccAceNorg_ADH_all")
cccAceNorgModel_ADH_all <- omxSetParameters( cccAceNorgModel_ADH_all, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )

cccAceNorgFit_ADH_all <- mxTryHard(cccAceNorgModel_ADH_all)
cccAceNorgFit_ADH_all_2 <- mxTryHard(cccAceNorgFit_ADH_all)
cccAceNorgFit_ADH_all_3 <- mxTryHard(cccAceNorgFit_ADH_all_2)
mxCompare(cccFitADH_all_3, cccAceNorgFit_ADH_all_3)
est_cccAceNorgFitADH_all <- summary(cccAceNorgFit_ADH_all_3)
write.csv(est_cccAceNorgFitADH_all$parameters, "est_cccAceNorgFitADH_all.csv")

# ----THm=THf----

cccAceNothModel_ADH_all <- mxModel(cccFitADH_all_3, name="cccAceNoth_ADH_all")
cccAceNothModel_ADH_all <- omxSetParameters( cccAceNothModel_ADH_all,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothModel_ADH_all <- omxSetParameters( cccAceNothModel_ADH_all,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothFit_ADH_all <- mxTryHard(cccAceNothModel_ADH_all)
cccAceNothFit_ADH_all_2 <- mxTryHard(cccAceNothFit_ADH_all)
cccAceNothFit_ADH_all_3 <- mxTryHard(cccAceNothFit_ADH_all_2)
cccAceNothFit_ADH_all_4 <- mxTryHard(cccAceNothFit_ADH_all_3)
est_cccAceNothFitADH_all <- summary(cccAceNothFit_ADH_all_2)
write.csv(est_cccAceNothFitADH_all$parameters, "est_cccAceNothFitADH_all.csv")

```

```

# ---- equal ACE estimate across sex ----

# equate rg, beta, and ACE across sex

cccAceModel_ADH_all <- mxModel(cccFitADH_all_3, name="cccAce_ADH_all")

cccAceModel_ADH_all <- omxSetParameters( cccAceModel_ADH_all, labels=c("rgi", "rgd","rgq"), free=FALSE, values=1 )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("am11","am21","am31", "am22","am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("cm11","cm21","cm31", "cm22","cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("em11","em21","em31", "em22","em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("af11","af21","af31", "af22","af32", "af33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("cf11","cf21","cf31", "cf22","cf32", "cf33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels=c("ef11","ef21","ef31", "ef22","ef32", "ef33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceModel_ADH_all <- omxSetParameters (cccAceModel_ADH_all, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceFit_ADH_all <- mxTryHard(cccAceModel_ADH_all)

cccAceFit_ADH_all_2 <- mxTryHard(cccAceFit_ADH_all)

cccAceFit_ADH_all_3 <- mxTryHard(cccAceFit_ADH_all_2)

cccAceFit_ADH_all_4 <- mxTryHard(cccAceFit_ADH_all_3)

.showFitStatistics(cccAceFit_ADH_all)

.showFitStatistics(cccAceFit_ADH_all_2)

.showFitStatistics(cccAceFit_ADH_all_3)

.showFitStatistics(cccAceFit_ADH_all_4)

est_cccAceFitADH_all <- summary(cccAceFit_ADH_all_4)

write.csv(est_cccAceFitADH_all$parameters, "est_cccAceFitADH_all.csv")

# ---- ACE with No beta ----

cccAceNoBModel_ADH_all <- mxModel(cccAceFit_ADH_all_4, name="cccAceNoB_ADH_all")

cccAceNoBModel_ADH_all <- omxSetParameters (cccAceNoBModel_ADH_all, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_ADH_all <- mxTryHard(cccAceNoBModel_ADH_all)

cccAceNoBFit_ADH_all_2 <- mxTryHard(cccAceNoBFit_ADH_all)

```

```

cccAceNoBFit_ADH_all_3 <- mxTryHard(cccAceNoBFit_ADH_all_2)
est_cccAceNoBFitADH_all <- summary(cccAceNoBFit_ADH_all_3)
write.csv(est_cccAceNoBFitADH_all$parameters, "est_cccAceNoBFitADH_all.csv")
# ---- ACE drop b_21 ----

cccAceNoB21Model_ADH_all <- mxModel(cccAceFit_ADH_all_4, name="cccAceNoB21_ADH_all")
cccAceNoB21Model_ADH_all <- omxSetParameters (cccAceNoB21Model_ADH_all, labels=c("b_21"), free = FALSE, values
= 0)
cccAceNoB21Fit_ADH_all <- mxTryHard(cccAceNoB21Model_ADH_all)
cccAceNoB21Fit_ADH_all_2 <- mxTryHard(cccAceNoB21Fit_ADH_all)
cccAceNoB21Fit_ADH_all_3 <- mxTryHard(cccAceNoB21Fit_ADH_all_2)
est_cccAceNo21BFitADH_all <- summary(cccAceNoB21Fit_ADH_all_3)
write.csv(est_cccAceNo21BFitADH_all$parameters, "est_cccAceNo21BFitADH_all.csv")
# ---- ACE drop b_31 ----

cccAceNoB31Model_ADH_all <- mxModel(cccAceFit_ADH_all_4, name="cccAceNoB31_ADH_all")
cccAceNoB31Model_ADH_all <- omxSetParameters (cccAceNoB31Model_ADH_all, labels=c("b_31"), free = FALSE, values
= 0)
cccAceNoB31Fit_ADH_all <- mxTryHard(cccAceNoB31Model_ADH_all)
cccAceNoB31Fit_ADH_all_2 <- mxTryHard(cccAceNoB31Fit_ADH_all)
cccAceNoB31Fit_ADH_all_3 <- mxTryHard(cccAceNoB31Fit_ADH_all_2)
cccAceNoB31Fit_ADH_all_4 <- mxTryHard(cccAceNoB31Fit_ADH_all_3)
est_cccAceNo31BFitADH_all <- summary(cccAceNoB31Fit_ADH_all_4)
write.csv(est_cccAceNo31BFitADH_all$parameters, "est_cccAceNo31BFitADH_all.csv")
# ---- ACE drop b31, drop C ----

cccAeNoB31Model_ADH_all <- mxModel(cccAceFit_ADH_all_4, name="cccAeNoB31_ADH_all")
cccAeNoB31Model_ADH_all <- omxSetParameters (cccAeNoB31Model_ADH_all, labels=c("b_31"), free = FALSE, values =
0)
cccAeNoB31Model_ADH_all <- omxSetParameters (cccAeNoB31Model_ADH_all, labels=c("c11", "c22", "c32", "c33"), free =
FALSE, values = 0)
cccAeNoB31Fit_ADH_all <- mxTryHard(cccAeNoB31Model_ADH_all)
cccAeNoB31Fit_ADH_all_2 <- mxTryHard(cccAeNoB31Fit_ADH_all)
cccAeNoB31Fit_ADH_all_3 <- mxTryHard(cccAeNoB31Fit_ADH_all_2)
est_cccAeNo31BFitADH_all <- summary(cccAeNoB31Fit_ADH_all_3)
write.csv(est_cccAeNo31BFitADH_all$parameters, "est_cccAeNo31BFitADH_all.csv")
# ---- ACE drop beta 31, drop A ----

```

```

cccCeNoB31Model_ADH_all <- mxModel(cccAceFit_ADH_all_4, name="cccCeNoB31_ADH_all")

cccCeNoB31Model_ADH_all <- omxSetParameters (cccCeNoB31Model_ADH_all, labels=c("b_31"), free = FALSE, values =
0)

cccCeNoB31Model_ADH_all <- omxSetParameters (cccCeNoB31Model_ADH_all, labels=c("a11", "a22", "a32", "a33"), free =
FALSE, values = 0)

cccCeNoB31Fit_ADH_all <- mxTryHard(cccCeNoB31Model_ADH_all)
cccCeNoB31Fit_ADH_all_2 <- mxTryHard(cccCeNoB31Fit_ADH_all)
cccCeNoB31Fit_ADH_all_3 <- mxTryHard(cccCeNoB31Fit_ADH_all_2)
est_cccCeNo31BFitADH_all <- summary(cccCeNoB31Fit_ADH_all_3)
write.csv(est_cccCeNo31BFitADH_all$parameters, "est_cccCeNo31BFitADH_all.csv")
# ---- All submodels ----
fitstatsubmodels1_ADH_all <- mxCompare(cccFitADH_all_3, nested <- list(cccAceNorgFit_ADH_all_3,
                                cccAceNothFit_ADH_all_2,
                                cccAceFit_ADH_all_4))
fitstatsubmodels2_ADH_all <- mxCompare(cccAceFit_ADH_all_4, nested <- list(cccAceNoBFit_ADH_all_3,
                                cccAceNoB21Fit_ADH_all_3,
                                cccAceNoB31Fit_ADH_all_4,
                                cccAeNoB31Fit_ADH_all_3,
                                cccCeNoB31Fit_ADH_all_3))
fitstatsubmodels_ADH_all <- rbind(fitstatsubmodels1_ADH_all, fitstatsubmodels2_ADH_all)
write.csv(fitstatsubmodels_ADH_all, "submodelsADH_all.csv")
# base model
write.csv(summary(cccFitADH_all_3)$parameters, "estimates_basemodel_ADH_all.csv")
# ----- Create Confidence Interval Objects: best fitting model -----
# run the best fitting with 95% CIs
bestci <- mxCI( c("a11", "a22", "a32", "a33",
                "e11", "e22", "e32", "e33",
                "b_21"))
bestmodel_ADH_all <- mxModel(cccAeNoB31Fit_ADH_all_3, bestci)
fitbest_ADH_all <- mxTryHard(bestmodel_ADH_all, intervals = F)
fitbest_ADH_all_2 <- mxTryHard(fitbest_ADH_all, intervals = F)
fitbest_ADH_all_3 <- mxTryHard(fitbest_ADH_all_2, intervals = F)
fitbest_ADH_all_int <- mxRun(fitbest_ADH_all_3, intervals = T)

```

```

sumbest_ADH_all <- summary(fitbest_ADH_all_int)

sumbest_ADH_all$parameters

sumbest_ADH_all$CI

sumbest_ADH_all$CI$CIint <- paste("(",round(sumbest_ADH_all$CI$lbound,2), "-",
round(sumbest_ADH_all$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_ADH_all$CI, "CI_bestfitting_ADH_all.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                      "c11", "c22", "c32", "c33",
                      "e11", "e22", "e32", "e33",
                      "b_21", "b_31"))

ACEmodel_ADH_all <- mxModel(cccAceFit_ADH_all_4, ACEciACEnorg)

fitACE_ADH_all <- mxTryHard(ACEmodel_ADH_all, intervals = F)

fitACE_ADH_all_2 <- mxTryHard(fitACE_ADH_all, intervals = F)

fitACE_ADH_all_3 <- mxTryHard(fitACE_ADH_all_2, intervals = F)

fitACE_ADH_all_int <- mxRun(fitACE_ADH_all_3, intervals = T)

sumACE_ADH_all <- summary(fitACE_ADH_all_int)

sumACE_ADH_all$parameters

sumACE_ADH_all$CI

sumACE_ADH_all$CI$CIint <- paste("(",round(sumACE_ADH_all$CI$lbound,2), "-",
round(sumACE_ADH_all$CI$subound,2),")", sep="")

sumACE_ADH_all1 <- sumACE_ADH_all$CI

sumACE_ADH_all1$datagroup <- "ADH_all"

write.csv(sumACE_ADH_all1, "sumACE_ADH_all.csv")

```

Appendix VI R script for CCC Model for MASATS age 12-15. (NCT: MASATS)

```
# -----Load Library-----

require(OpenMx)

require(psych)

source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")

library('polycor')

mxOption( NULL, "Default optimizer", "NPSOL" )

# -----

# ----- PREPARE DATA -----

data <- read.table("NCTGroup1215.csv", header=TRUE, sep=",")

twindat <- function(dat, famid, twinid, zygosity) {

  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1

  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2

  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))

  DAT[,paste(twinid, "_T1", sep="")] <- NULL

  DAT[,paste(twinid, "_T2", sep="")] <- NULL

  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])

  DAT[,paste(zygosity, "_T1", sep="")] <- NULL

  DAT[,paste(zygosity, "_T2", sep="")] <- NULL

  return(DAT)

}

# three arguments needed for twindat() function

nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zyggroup")

# Load Data

describe(nic, skew=F)

# Select Variables for Analysis

Vars <- c('tobeveer', 'maxcigxdayCCC', 'quiteverCCC')

nv <- 3 # number of variables

ntv <- nv*2 # number of total variables

selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")

# Specify Thresholds for Ordinal Variables
```



```

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1

nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))

thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)

thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

```

```

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )
covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )
covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )
covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )
covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)
nvI <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )
nvZ <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2 <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1 <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables
Var1m <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds
Mean <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

```

```

expCovMZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression= (rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf)))) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(am))+cm%*%t(cf)),
cbind(0.5%x%(rg*(af%*%t(am))+cf%*%t(cm)) , Af+Cf+Ef)), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo" )

funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf )
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm )
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )

```

```

modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )

modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelNCT1215 <- mxModel( "ccc_NCT1215", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ---- RUN GENETIC MODEL ----
# Run CCC and Cholesky Decomposition ACE model

cccFitNCT1215 <- mxTryHard(cccModelNCT1215, intervals=F)
cccFitNCT1215_2 <- mxTryHard(cccFitNCT1215, intervals=F)
cccFitNCT1215_3 <- mxTryHard(cccFitNCT1215_2, intervals=F)
est_cccFitNCT1215 <- summary(cccFitNCT1215_3)
write.csv(est_cccFitNCT1215$parameters, "est_cccFitNCT1215.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----

cccAceNorgModel_NCT1215 <- mxModel(cccFitNCT1215_3, name="cccAceNorg_NCT1215")

cccAceNorgModel_NCT1215 <- omxSetParameters( cccAceNorgModel_NCT1215, labels=c("rgi", "rgd", "rgq"), free=FALSE,
values=1 )

cccAceNorgFit_NCT1215 <- mxTryHard(cccAceNorgModel_NCT1215)
cccAceNorgFit_NCT1215_2 <- mxTryHard(cccAceNorgFit_NCT1215)
cccAceNorgFit_NCT1215_3 <- mxTryHard(cccAceNorgFit_NCT1215_2)
est_cccAceNorgFitNCT1215 <- summary(cccAceNorgFit_NCT1215_3)
write.csv(est_cccAceNorgFitNCT1215$parameters, "est_cccAceNorgFitNCT1215.csv")

# ----THm=THf----

cccAceNothModel_NCT1215 <- mxModel(cccFitNCT1215_3, name="cccAceNoth_NCT1215")
cccAceNothModel_NCT1215 <- omxSetParameters( cccAceNothModel_NCT1215,
labels=c("t1thmim", "t1thmdm", "t1thmqm", "t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))
cccAceNothModel_NCT1215 <- omxSetParameters( cccAceNothModel_NCT1215,
labels=c("t1thmif", "t1thmdf", "t1thmqf", "t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))

```

```

cccAceNothFit_NCT1215 <- mxTryHard(cccAceNothModel_NCT1215)

cccAceNothFit_NCT1215_2 <- mxTryHard(cccAceNothFit_NCT1215)

cccAceNothFit_NCT1215_3 <- mxTryHard(cccAceNothFit_NCT1215_2)

est_cccAceNothFitNCT1215 <- summary(cccAceNothFit_NCT1215_3)

write.csv(est_cccAceNothFitNCT1215$parameters, "est_cccAceNothFitNCT1215.csv")

# ---- equal ACE estimate across sex ----

# equate rg, beta, and ACE across sex

cccAceModel_NCT1215 <- mxModel(cccFitNCT1215_3, name="cccAce_NCT1215")

cccAceModel_NCT1215 <- omxSetParameters( cccAceModel_NCT1215, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("af11", "af21", "af31", "af22", "af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("cf11", "cf21", "cf31", "cf22", "cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels=c("ef11", "ef21", "ef31", "ef22", "ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceModel_NCT1215 <- omxSetParameters (cccAceModel_NCT1215, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceFit_NCT1215 <- mxTryHard(cccAceModel_NCT1215)

cccAceFit_NCT1215_2 <- mxTryHard(cccAceFit_NCT1215)

cccAceFit_NCT1215_3 <- mxTryHard(cccAceFit_NCT1215_2)

est_cccAceFitNCT1215 <- summary(cccAceFit_NCT1215_3)

write.csv(est_cccAceFitNCT1215$parameters, "est_cccAceFitNCT1215.csv")

# ---- ACE with No beta ----

cccAceNoBModel_NCT1215 <- mxModel(cccAceFit_NCT1215_3, name="cccAceNoB_NCT1215")

cccAceNoBModel_NCT1215 <- omxSetParameters (cccAceNoBModel_NCT1215, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_NCT1215 <- mxTryHard(cccAceNoBModel_NCT1215)

cccAceNoBFit_NCT1215_2 <- mxTryHard(cccAceNoBFit_NCT1215)

```

```

cccAceNoBFit_NCT1215_3 <- mxTryHard(cccAceNoBFit_NCT1215_2)
est_cccAceNoBFitNCT1215 <- summary(cccAceNoBFit_NCT1215_3)
write.csv(est_cccAceNoBFitNCT1215$parameters, "est_cccAceNoBFitNCT1215.csv")
# ---- ACE drop b_21 ----

cccAceNoB21Model_NCT1215 <- mxModel(cccAceFit_NCT1215_3, name="cccAceNoB21_NCT1215")
cccAceNoB21Model_NCT1215 <- omxSetParameters (cccAceNoB21Model_NCT1215, labels=c("b_21"), free = FALSE, values
= 0)
cccAceNoB21Fit_NCT1215 <- mxTryHard(cccAceNoB21Model_NCT1215)
cccAceNoB21Fit_NCT1215_2 <- mxTryHard(cccAceNoB21Fit_NCT1215)
cccAceNoB21Fit_NCT1215_3 <- mxTryHard(cccAceNoB21Fit_NCT1215_2)
est_cccAceNo21BFitNCT1215 <- summary(cccAceNoB21Fit_NCT1215_3)
write.csv(est_cccAceNo21BFitNCT1215$parameters, "est_cccAceNo21BFitNCT1215.csv")
# ---- ACE drop b_31 ----

cccAceNoB31Model_NCT1215 <- mxModel(cccAceFit_NCT1215_3, name="cccAceNoB31_NCT1215")
cccAceNoB31Model_NCT1215 <- omxSetParameters (cccAceNoB31Model_NCT1215, labels=c("b_31"), free = FALSE, values
= 0)
cccAceNoB31Fit_NCT1215 <- mxTryHard(cccAceNoB31Model_NCT1215)
cccAceNoB31Fit_NCT1215_2 <- mxTryHard(cccAceNoB31Fit_NCT1215)
cccAceNoB31Fit_NCT1215_3 <- mxTryHard(cccAceNoB31Fit_NCT1215_2)
est_cccAceNo31BFitNCT1215 <- summary(cccAceNoB31Fit_NCT1215_3)
write.csv(est_cccAceNo31BFitNCT1215$parameters, "est_cccAceNo31BFitNCT1215.csv")
# ---- ACE drop b31, drop C ----

cccAeNoB31Model_NCT1215 <- mxModel(cccAceFit_NCT1215_3, name="cccAeNoB31_NCT1215")
cccAeNoB31Model_NCT1215 <- omxSetParameters (cccAeNoB31Model_NCT1215, labels=c("b_31"), free = FALSE, values =
0)
cccAeNoB31Model_NCT1215 <- omxSetParameters (cccAeNoB31Model_NCT1215, labels=c("c11", "c22", "c32", "c33"), free
= FALSE, values = 0)
cccAeNoB31Fit_NCT1215 <- mxTryHard(cccAeNoB31Model_NCT1215)
cccAeNoB31Fit_NCT1215_2 <- mxTryHard(cccAeNoB31Fit_NCT1215)
cccAeNoB31Fit_NCT1215_3 <- mxTryHard(cccAeNoB31Fit_NCT1215_2)
est_cccAeNo31BFitNCT1215 <- summary(cccAeNoB31Fit_NCT1215_3)
write.csv(est_cccAeNo31BFitNCT1215$parameters, "est_cccAeNo31BFitNCT1215.csv")
# ---- ACE drop beta 31, drop A ----

cccCeNoB31Model_NCT1215 <- mxModel(cccAceFit_NCT1215_3, name="cccCeNoB31_NCT1215")

```

```

cccCeNoB31Model_NCT1215 <- omxSetParameters (cccCeNoB31Model_NCT1215, labels=c("b_31"), free = FALSE, values =
0)

cccCeNoB31Model_NCT1215 <- omxSetParameters (cccCeNoB31Model_NCT1215, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)

cccCeNoB31Fit_NCT1215 <- mxTryHard(cccCeNoB31Model_NCT1215)
cccCeNoB31Fit_NCT1215_2 <- mxTryHard(cccCeNoB31Fit_NCT1215)
cccCeNoB31Fit_NCT1215_3 <- mxTryHard(cccCeNoB31Fit_NCT1215_2)
est_cccCeNo31BFitNCT1215 <- summary(cccCeNoB31Fit_NCT1215_3)
write.csv(est_cccCeNo31BFitNCT1215$parameters, "est_cccCeNo31BFitNCT1215.csv")

# ---- All submodels ----

fitstatsubmodels1_NCT1215 <- mxCompare(cccFitNCT1215_3, nested <- list(cccAceNorgFit_NCT1215_3,
                                cccAceNothFit_NCT1215_3,
                                cccAceFit_NCT1215_3))

fitstatsubmodels2_NCT1215 <- mxCompare(cccAceFit_NCT1215_3, nested <- list(cccAceNoBFit_NCT1215_3,
                                cccAceNoB21Fit_NCT1215_3,
                                cccAceNoB31Fit_NCT1215_3,
                                cccAeNoB31Fit_NCT1215_3,
                                cccCeNoB31Fit_NCT1215_3))

fitstatsubmodels_NCT1215 <- rbind(fitstatsubmodels1_NCT1215, fitstatsubmodels2_NCT1215)
write.csv(fitstatsubmodels_NCT1215, "submodelsNCT1215.csv")

# base model

write.csv(summary(cccFitNCT1215_3)$parameters, "estimates_basemodel_NCT_1215.csv")

# ---- Create Confidence Interval Objects: best fitting model ----

# run the best fitting with 95% CIs

bestci <- mxCI( c("c11", "c22", "c32", "c33",
                 "e11", "e22", "e32", "e33",
                 "b_21"))

bestmodel_NCT_1215 <- mxModel(cccCeNoB31Fit_NCT1215_3, bestci)
fitbest_NCT_1215 <- mxTryHard(bestmodel_NCT_1215, intervals = F)
fitbest_NCT_1215_2 <- mxTryHard(fitbest_NCT_1215, intervals = F)
fitbest_NCT_1215_3 <- mxTryHard(fitbest_NCT_1215_2, intervals = F)
fitbest_NCT_1215_int <- mxRun(fitbest_NCT_1215_3, intervals = T)
sumbest_NCT_1215 <- summary(fitbest_NCT_1215_int)

```



```

sumbest_NCT_1215$parameters

sumbest_NCT_1215$CI

sumbest_NCT_1215$CI$CIint <- paste("(",round(sumbest_NCT_1215$CI$lbound,2), "-",
round(sumbest_NCT_1215$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_NCT_1215$CI, "CI_bestfitting_NCT_1215.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                        "c11", "c22", "c32", "c33",
                        "e11", "e22", "e32", "e33",
                        "b_21", "b_31"))

ACEmodel_NCT_1215 <- mxModel(cccAceFit_NCT1215_3, ACEciACEnorg)

fitACE_NCT_1215 <- mxTryHard(ACEmodel_NCT_1215, intervals = F)

fitACE_NCT_1215_2 <- mxTryHard(fitACE_NCT_1215, intervals = F)

fitACE_NCT_1215_3 <- mxTryHard(fitACE_NCT_1215_2, intervals = F)

fitACE_NCT_1215_int <- mxRun(fitACE_NCT_1215_3, intervals = T)

sumACE_NCT_1215 <- summary(fitACE_NCT_1215_int)

sumACE_NCT_1215$parameters

sumACE_NCT_1215$CI

sumACE_NCT_1215$CI$CIint <- paste("(",round(sumACE_NCT_1215$CI$lbound,2), "-",
round(sumACE_NCT_1215$CI$subound,2),")", sep="")

sumACE_NCT_12151 <- sumACE_NCT_1215$CI

sumACE_NCT_12151$datagroup <- "NCT_1215"

write.csv(sumACE_NCT_12151, "sumACE_NCT_1215.csv")

```

Appendix VII R script for CCC Model for MASATS age 16-19. (NCT: MASATS)

```
# -----Load Library-----
require(OpenMx)
require(psych)
source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )
# -----
# ----- PREPARE DATA -----
data <- read.table("NCTGroup1619.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobever', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```

```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1 <- 1 ; fcat1 <- 0 ; lcat1 <- fcat1+nth1 ; ncat1 <- nth1+1

nth2 <- 2 ; fcat2 <- 0 ; lcat2 <- fcat2+nth2 ; ncat2 <- nth2+1

nth3 <- 1 ; fcat3 <- 0 ; lcat3 <- fcat3+nth3 ; ncat3 <- nth3+1

nth <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1 <- 0 ; ith1 <- 0.1 ;

lth2 <- 0 ; ith2 <- 0.1 ;

lth3 <- 0 ; ith3 <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))

thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)

thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))

twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))

twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)

dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)

dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)

dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)

dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----

# -----

```

```

# PREPARE GENETIC MODEL

# -----

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )

covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )

covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )

covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )

covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

```

```

# Algebra to compute total variances and standard deviations (diagonal only)
nvI    <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )
nvZ    <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2     <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1   <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm     <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf     <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm   <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf   <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables
Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds
Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )
expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix
expCovMZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

```

```

expCovDZo <- mxAlgebra( expression=( rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf))) )&&%
                                rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af))+cm%*%t(cf)),
                                cbind(0.5%x%(rg*(af%*%t(am))+cf%*%t(cm) , Af+Cf+Ef)), name="expCovDZo" )

# Data objects for Multiple Groups

dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups

expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )

expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )

expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )

expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )

expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo" )

funML <- mxFitFunctionML()

# Combine Groups

pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )

parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)

parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)

modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelNCT1619 <- mxModel( "ccc_NCT1619", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----

```

```

# ---- RUN GENETIC MODEL ----

# Run CCC and Cholesky Decomposition ACE model

cccFitNCT1619 <- mxTryHard(cccModelNCT1619, intervals=F)
cccFitNCT1619_2 <- mxTryHard(cccFitNCT1619, intervals=F)
cccFitNCT1619_3 <- mxTryHard(cccFitNCT1619_2, intervals=F)
est_cccFitNCT1619 <- summary(cccFitNCT1619_3)
write.csv(est_cccFitNCT1619$parameters, "est_cccFitNCT1619.csv")

# Test Submodels constraining sex

# -----
# ---- No rg ----

cccAceNorgModel_NCT1619 <- mxModel(cccFitNCT1619_3, name="cccAceNorg_NCT1619")

cccAceNorgModel_NCT1619 <- omxSetParameters( cccAceNorgModel_NCT1619, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )

cccAceNorgFit_NCT1619 <- mxTryHard(cccAceNorgModel_NCT1619)
cccAceNorgFit_NCT1619_2 <- mxTryHard(cccAceNorgFit_NCT1619)
cccAceNorgFit_NCT1619_3 <- mxTryHard(cccAceNorgFit_NCT1619_2)
mxCompare(cccFitNCT1619_3, cccAceNorgFit_NCT1619_3)
est_cccAceNorgFitNCT1619 <- summary(cccAceNorgFit_NCT1619_3)
write.csv(est_cccAceNorgFitNCT1619$parameters, "est_cccAceNorgFitNCT1619.csv")

# ----THm=THf-----

cccAceNothModel_NCT1619 <- mxModel(cccFitNCT1619_3, name="cccAceNoth_NCT1619")

cccAceNothModel_NCT1619 <- omxSetParameters( cccAceNothModel_NCT1619,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothModel_NCT1619 <- omxSetParameters( cccAceNothModel_NCT1619,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothFit_NCT1619 <- mxTryHard(cccAceNothModel_NCT1619)
cccAceNothFit_NCT1619_2 <- mxTryHard(cccAceNothFit_NCT1619)
est_cccAceNothFitNCT1619 <- summary(cccAceNothFit_NCT1619_2)
write.csv(est_cccAceNothFitNCT1619$parameters, "est_cccAceNothFitNCT1619.csv")

# ---- equal ACE estimate across sex ----

```

```

# equate rg, beta, and ACE across sex

cccAceModel_NCT1619 <- mxModel(cccFitNCT1619_3, name="cccAce_NCT1619")

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("af11", "af21", "af31", "af22", "af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("cf11", "cf21", "cf31", "cf22", "cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels=c("ef11", "ef21", "ef31", "ef22", "ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceModel_NCT1619 <- omxSetParameters( cccAceModel_NCT1619, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))

cccAceFit_NCT1619 <- mxTryHard(cccAceModel_NCT1619)

cccAceFit_NCT1619_2 <- mxTryHard(cccAceFit_NCT1619)

cccAceFit_NCT1619_3 <- mxTryHard(cccAceFit_NCT1619_2)

cccAceFit_NCT1619_4 <- mxTryHard(cccAceFit_NCT1619_3)

cccAceFit_NCT1619_5 <- mxTryHard(cccAceFit_NCT1619_4)

est_cccAceFitNCT1619 <- summary(cccAceFit_NCT1619_3)

write.csv(est_cccAceFitNCT1619$parameters, "est_cccAceFitNCT1619.csv")

# ---- ACE with No beta ----

cccAceNoBModel_NCT1619 <- mxModel(cccAceFit_NCT1619_3, name="cccAceNoB_NCT1619")

cccAceNoBModel_NCT1619 <- omxSetParameters( cccAceNoBModel_NCT1619, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_NCT1619 <- mxTryHard(cccAceNoBModel_NCT1619)

cccAceNoBFit_NCT1619_2 <- mxTryHard(cccAceNoBFit_NCT1619)

cccAceNoBFit_NCT1619_3 <- mxTryHard(cccAceNoBFit_NCT1619_2)

cccAceNoBFit_NCT1619_4 <- mxTryHard(cccAceNoBFit_NCT1619_3)

est_cccAceNoBFitNCT1619 <- summary(cccAceNoBFit_NCT1619_4)

write.csv(est_cccAceNoBFitNCT1619$parameters, "est_cccAceNoBFitNCT1619.csv")

```



```

# ---- ACE drop b_21 ----
cccAceNoB21Model_NCT1619 <- mxModel(cccAceFit_NCT1619_3, name="cccAceNoB21_NCT1619")
cccAceNoB21Model_NCT1619 <- omxSetParameters (cccAceNoB21Model_NCT1619, labels=c("b_21"), free = FALSE, values
= 0)
cccAceNoB21Fit_NCT1619 <- mxTryHard(cccAceNoB21Model_NCT1619)
cccAceNoB21Fit_NCT1619_2 <- mxTryHard(cccAceNoB21Fit_NCT1619)
cccAceNoB21Fit_NCT1619_3 <- mxTryHard(cccAceNoB21Fit_NCT1619_2)
est_cccAceNo21BFitNCT1619 <- summary(cccAceNoB21Fit_NCT1619_3)
write.csv(est_cccAceNo21BFitNCT1619$parameters, "est_cccAceNo21BFitNCT1619.csv")

# ---- ACE drop b_31 ----
cccAceNoB31Model_NCT1619 <- mxModel(cccAceFit_NCT1619_3, name="cccAceNoB31_NCT1619")
cccAceNoB31Model_NCT1619 <- omxSetParameters (cccAceNoB31Model_NCT1619, labels=c("b_31"), free = FALSE, values
= 0)
cccAceNoB31Fit_NCT1619 <- mxTryHard(cccAceNoB31Model_NCT1619)
cccAceNoB31Fit_NCT1619_2 <- mxTryHard(cccAceNoB31Fit_NCT1619)
cccAceNoB31Fit_NCT1619_3 <- mxTryHard(cccAceNoB31Fit_NCT1619_2)
cccAceNoB31Fit_NCT1619_4 <- mxTryHard(cccAceNoB31Fit_NCT1619_3)
cccAceNoB31Fit_NCT1619_5 <- mxTryHard(cccAceNoB31Fit_NCT1619_4)
est_cccAceNo31BFitNCT1619 <- summary(cccAceNoB31Fit_NCT1619)
write.csv(est_cccAceNo31BFitNCT1619$parameters, "est_cccAceNo31BFitNCT1619.csv")

# ---- ACE drop b31, drop C ----
cccAeNoB31Model_NCT1619 <- mxModel(cccAceFit_NCT1619_3, name="cccAeNoB31_NCT1619")
cccAeNoB31Model_NCT1619 <- omxSetParameters (cccAeNoB31Model_NCT1619, labels=c("b_31"), free = FALSE, values =
0)
cccAeNoB31Model_NCT1619 <- omxSetParameters (cccAeNoB31Model_NCT1619, labels=c("c11", "c22", "c32", "c33"), free
= FALSE, values = 0)
cccAeNoB31Fit_NCT1619 <- mxTryHard(cccAeNoB31Model_NCT1619)
cccAeNoB31Fit_NCT1619_2 <- mxTryHard(cccAeNoB31Fit_NCT1619)
cccAeNoB31Fit_NCT1619_3 <- mxTryHard(cccAeNoB31Fit_NCT1619_2)
cccAeNoB31Fit_NCT1619_4 <- mxTryHard(cccAeNoB31Fit_NCT1619_3)
cccAeNoB31Fit_NCT1619_5 <- mxTryHard(cccAeNoB31Fit_NCT1619_4)
cccAeNoB31Fit_NCT1619_6 <- mxTryHard(cccAeNoB31Fit_NCT1619_5)
est_cccAeNo31BFitNCT1619 <- summary(cccAeNoB31Fit_NCT1619_6)
write.csv(est_cccAeNo31BFitNCT1619$parameters, "est_cccAeNo31BFitNCT1619.csv")

```

```

# ---- ACE drop beta 31, drop A ----
cccCeNoB31Model_NCT1619 <- mxModel(cccAceFit_NCT1619_3, name="cccCeNoB31_NCT1619")
cccCeNoB31Model_NCT1619 <- omxSetParameters (cccCeNoB31Model_NCT1619, labels=c("b_31"), free = FALSE, values =
0)
cccCeNoB31Model_NCT1619 <- omxSetParameters (cccCeNoB31Model_NCT1619, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)
cccCeNoB31Fit_NCT1619 <- mxTryHard(cccCeNoB31Model_NCT1619)
cccCeNoB31Fit_NCT1619_2 <- mxTryHard(cccCeNoB31Fit_NCT1619)
cccCeNoB31Fit_NCT1619_3 <- mxTryHard(cccCeNoB31Fit_NCT1619_2)
est_cccCeNo31BFitNCT1619 <- summary(cccCeNoB31Fit_NCT1619_3)
write.csv(est_cccCeNo31BFitNCT1619$parameters, "est_cccCeNo31BFitNCT1619.csv")
# ---- All submodels ----
fitstatsubmodels1_NCT1619 <- mxCompare(cccFitNCT1619_3, nested <- list(cccAceNorgFit_NCT1619_3,
cccAceNothFit_NCT1619_2,
cccAceFit_NCT1619_3))
fitstatsubmodels2_NCT1619 <- mxCompare(cccAceFit_NCT1619_3, nested <- list(cccAceNoBFit_NCT1619_4,
cccAceNoB21Fit_NCT1619_3,
cccAceNoB31Fit_NCT1619_5,
cccAeNoB31Fit_NCT1619_6,
cccCeNoB31Fit_NCT1619_3))
fitstatsubmodels_NCT1619 <- rbind(fitstatsubmodels1_NCT1619, fitstatsubmodels2_NCT1619)
write.csv(fitstatsubmodels_NCT1619, "submodelsNCT1619.csv")
# base model
write.csv(summary(cccFitNCT1619_3)$parameters, "estimates_basemodel_NCT_1619.csv")
# ---- Create Confidence Interval Objects: best fitting model ----
# run the best fitting with 95% CIs
bestci <- mxCI( c("a11", "a22", "a32", "a33",
"e11", "e22", "e32", "e33",
"b_21"))
bestmodel_NCT_1619 <- mxModel(cccAeNoB31Fit_NCT1619_6, bestci)
fitbest_NCT_1619 <- mxTryHard(bestmodel_NCT_1619, intervals = F)
fitbest_NCT_1619_2 <- mxTryHard(fitbest_NCT_1619, intervals = F)
fitbest_NCT_1619_3 <- mxTryHard(fitbest_NCT_1619_2, intervals = F)

```

```

fitbest_NCT_1619_int <- mxRun(fitbest_NCT_1619_3, intervals = T)

sumbest_NCT_1619 <- summary(fitbest_NCT_1619_int)

sumbest_NCT_1619$parameters

sumbest_NCT_1619$CI

sumbest_NCT_1619$CI$CIint <- paste("(",round(sumbest_NCT_1619$CI$lbound,2), "-",
round(sumbest_NCT_1619$CI$subound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_NCT_1619$CI, "CI_bestfitting_NCT_1619.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                        "c11", "c22", "c32", "c33",
                        "e11", "e22", "e32", "e33",
                        "b_21", "b_31"))

ACEmodel_NCT_1619 <- mxModel(cccAceFit_NCT1619_3, ACEciACEnorg)

fitACE_NCT_1619 <- mxTryHard(ACEmodel_NCT_1619, intervals = F)

fitACE_NCT_1619_2 <- mxTryHard(fitACE_NCT_1619, intervals = F)

fitACE_NCT_1619_3 <- mxTryHard(fitACE_NCT_1619_2, intervals = F)

fitACE_NCT_1619_int <- mxRun(fitACE_NCT_1619_3, intervals = T)

sumACE_NCT_1619 <- summary(fitACE_NCT_1619_int)

sumACE_NCT_1619$parameters

sumACE_NCT_1619$CI

sumACE_NCT_1619$CI$CIint <- paste("(",round(sumACE_NCT_1619$CI$lbound,2), "-",
round(sumACE_NCT_1619$CI$subound,2),")", sep="")

sumACE_NCT_16191 <- sumACE_NCT_1619$CI

sumACE_NCT_16191$datagroup <- "NCT_1619"

write.csv(sumACE_NCT_16191, "sumACE_NCT_1619.csv")

```

Appendix VIII R script for CCC Model for MASATS all participants. (NCT: MASATS)

```
# -----Load Library-----
require(OpenMx)
require(psych)
source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )
# -----
# ----- PREPARE DATA -----
data <- read.table("NCTbuild.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobever', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```

```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1 <- 1 ; fcat1 <- 0 ; lcat1 <- fcat1+nth1 ; ncat1 <- nth1+1
nth2 <- 2 ; fcat2 <- 0 ; lcat2 <- fcat2+nth2 ; ncat2 <- nth2+1
nth3 <- 1 ; fcat3 <- 0 ; lcat3 <- fcat3+nth3 ; ncat3 <- nth3+1
nth <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1 <- 0 ; ith1 <- 0.1 ;
lth2 <- 0 ; ith2 <- 0.1 ;
lth3 <- 0 ; ith3 <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))
thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)
thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]
describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0,
0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T),
                    values=c(0.6,0,0, 0.6, 0.6, 0.6), labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-
1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

```

```

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )
covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )
covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )
covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )
covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)
nvI <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )
nvZ <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2 <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1 <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables
Var1m <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds
Mean <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

```

```

expCovMZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression= (rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf)))) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(am))+cm%*%t(cf)),
cbind(0.5%x%(rg*(af%*%t(am))+cf%*%t(cm)) , Af+Cf+Ef)), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo" )

funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )

```



```

modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )

modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelNCT_all <- mxModel( "ccc_NCT_all", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ---- RUN GENETIC MODEL ----

# Run CCC and Cholesky Decomposition ACE model

cccFitNCT_all <- mxTryHard(cccModelNCT_all, intervals=F)
cccFitNCT_all_2 <- mxTryHard(cccFitNCT_all, intervals=F)
cccFitNCT_all_3 <- mxTryHard(cccFitNCT_all_2, intervals=F)
cccFitNCT_all_4 <- mxTryHard(cccFitNCT_all_3, intervals=F)
est_cccFitNCT_all <- summary(cccFitNCT_all_4)

write.csv(est_cccFitNCT_all$parameters, "est_cccFitNCT_all.csv")

# Test Submodels constraining sex

# -----
# ---- No rg ----

cccAceNorgModel_NCT_all <- mxModel(cccFitNCT_all_4, name="cccAceNorg_NCT_all")

cccAceNorgModel_NCT_all <- omxSetParameters( cccAceNorgModel_NCT_all, labels=c("rgi", "rgd", "rgq"), free=FALSE,
values=1 )

cccAceNorgFit_NCT_all <- mxTryHard(cccAceNorgModel_NCT_all)
cccAceNorgFit_NCT_all_2 <- mxTryHard(cccAceNorgFit_NCT_all)
cccAceNorgFit_NCT_all_3 <- mxTryHard(cccAceNorgFit_NCT_all_2)
est_cccAceNorgFitNCT_all <- summary(cccAceNorgFit_NCT_all_3)

write.csv(est_cccAceNorgFitNCT_all$parameters, "est_cccAceNorgFitNCT_all.csv")

# ---- THm=THf----

cccAceNothModel_NCT_all <- mxModel(cccFitNCT_all_4, name="cccAceNoth_NCT_all")
cccAceNothModel_NCT_all <- omxSetParameters( cccAceNothModel_NCT_all,
labels=c("t1thmim", "t1thmdm", "t1thmqm", "t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))

cccAceNothModel_NCT_all <- omxSetParameters( cccAceNothModel_NCT_all,

```

```

labels=c("t1thmif", "t1thmdf", "t1thmqf", "t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))
cccAceNothFit_NCT_all <- mxTryHard(cccAceNothModel_NCT_all)
cccAceNothFit_NCT_all_2 <- mxTryHard(cccAceNothFit_NCT_all)
cccAceNothFit_NCT_all_3 <- mxTryHard(cccAceNothFit_NCT_all_2)
cccAceNothFit_NCT_all_4 <- mxTryHard(cccAceNothFit_NCT_all_3)
est_cccAceNothFitNCT_all <- summary(cccAceNothFit_NCT_all)
write.csv(est_cccAceNothFitNCT_all$parameters, "est_cccAceNothFitNCT_all.csv")
# ---- equal ACE estimate across sex ----
# equate rg, beta, and ACE across sex
cccAceModel_NCT_all <- mxModel(cccFitNCT_all_3, name="cccAce_NCT_all")
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("af11", "af21", "af31", "af22", "af32", "af33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("cf11", "cf21", "cf31", "cf22", "cf32", "cf33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels=c("ef11", "ef21", "ef31", "ef22", "ef32", "ef33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))
cccAceModel_NCT_all <- omxSetParameters( cccAceModel_NCT_all, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))
cccAceFit_NCT_all <- mxTryHard(cccAceModel_NCT_all)
cccAceFit_NCT_all_2 <- mxTryHard(cccAceFit_NCT_all)
cccAceFit_NCT_all_3 <- mxTryHard(cccAceFit_NCT_all_2)
cccAceFit_NCT_all_4 <- mxTryHard(cccAceFit_NCT_all_3)
cccAceFit_NCT_all_5 <- mxTryHard(cccAceFit_NCT_all_4)
est_cccAceFitNCT_all <- summary(cccAceFit_NCT_all_5)
write.csv(est_cccAceFitNCT_all$parameters, "est_cccAceFitNCT_all.csv")
# ---- ACE with No beta ----

```

```

cccAceNoBModel_NCT_all <- mxModel(cccAceFit_NCT_all_5, name="cccAceNoB_NCT_all")

cccAceNoBModel_NCT_all <- omxSetParameters (cccAceNoBModel_NCT_all, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_NCT_all <- mxTryHard(cccAceNoBModel_NCT_all)

cccAceNoBFit_NCT_all_2 <- mxTryHard(cccAceNoBFit_NCT_all)

cccAceNoBFit_NCT_all_3 <- mxTryHard(cccAceNoBFit_NCT_all_2)

cccAceNoBFit_NCT_all_4 <- mxTryHard(cccAceNoBFit_NCT_all_3)

est_cccAceNoBFitNCT_all <- summary(cccAceNoBFit_NCT_all_4)

write.csv(est_cccAceNoBFitNCT_all$parameters, "est_cccAceNoBFitNCT_all.csv")

# ---- ACE drop b_21 ----

cccAceNoB21Model_NCT_all <- mxModel(cccAceFit_NCT_all_5, name="cccAceNoB21_NCT_all")

cccAceNoB21Model_NCT_all <- omxSetParameters (cccAceNoB21Model_NCT_all, labels=c("b_21"), free = FALSE, values =
0)

cccAceNoB21Fit_NCT_all <- mxTryHard(cccAceNoB21Model_NCT_all)

cccAceNoB21Fit_NCT_all_2 <- mxTryHard(cccAceNoB21Fit_NCT_all)

cccAceNoB21Fit_NCT_all_3 <- mxTryHard(cccAceNoB21Fit_NCT_all_2)

est_cccAceNo21BFitNCT_all <- summary(cccAceNoB21Fit_NCT_all_3)

write.csv(est_cccAceNo21BFitNCT_all$parameters, "est_cccAceNo21BFitNCT_all.csv")

# ---- ACE drop b_31 ----

cccAceNoB31Model_NCT_all <- mxModel(cccAceFit_NCT_all_5, name="cccAceNoB31_NCT_all")

cccAceNoB31Model_NCT_all <- omxSetParameters (cccAceNoB31Model_NCT_all, labels=c("b_31"), free = FALSE, values =
0)

cccAceNoB31Fit_NCT_all <- mxTryHard(cccAceNoB31Model_NCT_all)

cccAceNoB31Fit_NCT_all_2 <- mxTryHard(cccAceNoB31Fit_NCT_all)

cccAceNoB31Fit_NCT_all_3 <- mxTryHard(cccAceNoB31Fit_NCT_all_2)

est_cccAceNo31BFitNCT_all <- summary(cccAceNoB31Fit_NCT_all_3)

write.csv(est_cccAceNo31BFitNCT_all$parameters, "est_cccAceNo31BFitNCT_all.csv")

# ---- ACE drop b31, drop C ----

cccAeNoB31Model_NCT_all <- mxModel(cccAceFit_NCT_all_5, name="cccAeNoB31_NCT_all")

cccAeNoB31Model_NCT_all <- omxSetParameters (cccAeNoB31Model_NCT_all, labels=c("b_31"), free = FALSE, values = 0)

cccAeNoB31Model_NCT_all <- omxSetParameters (cccAeNoB31Model_NCT_all, labels=c("c11", "c22", "c32", "c33"), free =
FALSE, values = 0)

cccAeNoB31Fit_NCT_all <- mxTryHard(cccAeNoB31Model_NCT_all)

```

```

cccAeNoB31Fit_NCT_all_2 <- mxTryHard(cccAeNoB31Fit_NCT_all)
cccAeNoB31Fit_NCT_all_3 <- mxTryHard(cccAeNoB31Fit_NCT_all_2)
cccAeNoB31Fit_NCT_all_4 <- mxTryHard(cccAeNoB31Fit_NCT_all_3)
cccAeNoB31Fit_NCT_all_5 <- mxTryHard(cccAeNoB31Fit_NCT_all_4)
est_cccAeNo31BFitNCT_all <- summary(cccAeNoB31Fit_NCT_all_5)
write.csv(est_cccAeNo31BFitNCT_all$parameters, "est_cccAeNo31BFitNCT_all.csv")
# ---- ACE drop beta 31, drop A ----
cccCeNoB31Model_NCT_all <- mxModel(cccAceFit_NCT_all_5, name="cccCeNoB31_NCT_all")
cccCeNoB31Model_NCT_all <- omxSetParameters (cccCeNoB31Model_NCT_all, labels=c("b_31"), free = FALSE, values = 0)
cccCeNoB31Model_NCT_all <- omxSetParameters (cccCeNoB31Model_NCT_all, labels=c("a11", "a22", "a32", "a33"), free =
FALSE, values = 0)
cccCeNoB31Fit_NCT_all <- mxTryHard(cccCeNoB31Model_NCT_all)
cccCeNoB31Fit_NCT_all_2 <- mxTryHard(cccCeNoB31Fit_NCT_all)
cccCeNoB31Fit_NCT_all_3 <- mxTryHard(cccCeNoB31Fit_NCT_all_2)
est_cccCeNo31BFitNCT_all <- summary(cccCeNoB31Fit_NCT_all_3)
write.csv(est_cccCeNo31BFitNCT_all$parameters, "est_cccCeNo31BFitNCT_all.csv")
# ---- All submodels ----
fitstatsubmodels1_NCT_all <- mxCompare(cccFitNCT_all_4, nested <- list(cccAceNorgFit_NCT_all_3,
                                cccAceNothFit_NCT_all,
                                cccAceFit_NCT_all_5))
fitstatsubmodels2_NCT_all <- mxCompare(cccAceFit_NCT_all_5, nested <- list(cccAceNoBFit_NCT_all_4,
                                cccAceNoB21Fit_NCT_all_3,
                                cccAceNoB31Fit_NCT_all_3,
                                cccAeNoB31Fit_NCT_all_5,
                                cccCeNoB31Fit_NCT_all_3))
fitstatsubmodels_NCT_all <- rbind(fitstatsubmodels1_NCT_all, fitstatsubmodels2_NCT_all)
write.csv(fitstatsubmodels_NCT_all, "submodelsNCT_all.csv")
# base model
write.csv(summary(cccAceNoB31Fit_NCT_all_3)$parameters, "estimates_basemodel_NCT_all.csv")
# ----- Create Confidence Interval Objects: best fitting model ----
# run the best fitting with 95% CIs
bestci <- mxCI( c("a11", "a22", "a32", "a33",

```

```

        "c11", "c22", "c32", "c33",
        "e11", "e22", "e32", "e33",
        "b_21"))
bestmodel_NCT_all <- mxModel(cccAceNoB31Fit_NCT_all_3, bestci)
fitbest_NCT_all <- mxTryHard(bestmodel_NCT_all, intervals = F)
fitbest_NCT_all_2 <- mxTryHard(fitbest_NCT_all, intervals = F)
fitbest_NCT_all_3 <- mxTryHard(fitbest_NCT_all_2, intervals = F)
fitbest_NCT_all_int <- mxRun(fitbest_NCT_all_3, intervals = T)
sumbest_NCT_all <- summary(fitbest_NCT_all_int)
sumbest_NCT_all$parameters
sumbest_NCT_all$CI
sumbest_NCT_all$CI$CIint <- paste("(",round(sumbest_NCT_all$CI$lbound,2), "-",
round(sumbest_NCT_all$CI$subound,2),")", sep="")
# CI of best fitting model
write.csv(sumbest_NCT_all$CI, "CI_bestfitting_NCT_all.csv")
# ----- Create Confidence Interval Objects: ACE model no sex dif-----
ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
        "c11", "c22", "c32", "c33",
        "e11", "e22", "e32", "e33",
        "b_21", "b_31"))
ACEmodel_NCT_all <- mxModel(cccAceFit_NCT_all_5, ACEciACEnorg)
fitACE_NCT_all <- mxTryHard(ACEmodel_NCT_all, intervals = F)
fitACE_NCT_all_2 <- mxTryHard(fitACE_NCT_all, intervals = F)
fitACE_NCT_all_3 <- mxTryHard(fitACE_NCT_all_2, intervals = F)
fitACE_NCT_all_int <- mxRun(fitACE_NCT_all_3, intervals = T)
sumACE_NCT_all <- summary(fitACE_NCT_all_int)
sumACE_NCT_all$parameters
sumACE_NCT_all$CI
sumACE_NCT_all$CI$CIint <- paste("(",round(sumACE_NCT_all$CI$lbound,2), "-",
round(sumACE_NCT_all$CI$subound,2),")", sep="")
sumACE_NCT_all1 <- sumACE_NCT_all$CI
sumACE_NCT_all1$datagroup <- "NCT_all"
write.csv(sumACE_NCT_all1, "sumACE_NCT_all.csv")

```

Appendix IX R script for CCC Model for VTSABD/TSA age 22-32. (ABD: VTSABD/TSA)

```
# -----Load Library-----
require(OpenMx)
require(psych)
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )

# -----
# ----- PREPARE DATA -----
data <- read.table("ABDGroupTSA.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobeaver', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
# Specify Thresholds for Ordinal Variables
```

```

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1

nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))

thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)

thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]

describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

# Heterogeneity CCC ACE Model

```

```

# -----
# Specify Arguments for Causal Path Matrices

svB    <- 0
freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")
labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )
covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )
covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )
covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )
covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )
covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvI <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvI" )
nvZ <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )

```



```

I2    <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1  <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm    <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf    <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm  <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf  <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables

Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds

Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

expCovMZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression= (I2 %x% solve(nvI-bm)) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression= (I2 %x% solve(nvI-bf)) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression= (rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf)))) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af)))+cm%*%t(cf)),
cbind(0.5%x%(rg*(af%*%t(am)))+cf%*%t(cm) , Af+Cf+Ef)), name="expCovDZo" )

```

```

# Data objects for Multiple Groups

dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups

expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )

expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )

expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf" )

expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm" )

expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo" )

funML <- mxFitFunctionML()

# Combine Groups

pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )

parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf )

parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm )

modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )

multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))

cccModelABD2232 <- mxModel( "ccc_ABD2232", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi )

# -----
# ----- RUN GENETIC MODEL -----

```

```

# Run CCC and Cholesky Decomposition ACE model

cccFitABD2232 <- mxTryHard(cccModelABD2232, intervals=F)
cccFitABD2232_2 <- mxTryHard(cccFitABD2232, intervals=F)
cccFitABD2232_3 <- mxTryHard(cccFitABD2232_2, intervals=F)
cccFitABD2232_4 <- mxTryHard(cccFitABD2232_3, intervals=F)
cccFitABD2232_5 <- mxTryHard(cccFitABD2232_4, intervals=F)
est_cccFitABD2232 <- summary(cccFitABD2232_3)

write.csv(est_cccFitABD2232$parameters, "est_cccFitABD2232.csv")

# Test Submodels constraining sex
# -----
# ---- No rg ----

cccAceNorgModel_ABD2232 <- mxModel(cccFitABD2232_3, name="cccAceNorg_ABD2232")

cccAceNorgModel_ABD2232 <- omxSetParameters( cccAceNorgModel_ABD2232, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )

cccAceNorgFit_ABD2232 <- mxTryHard(cccAceNorgModel_ABD2232)
cccAceNorgFit_ABD2232_2 <- mxTryHard(cccAceNorgFit_ABD2232)
cccAceNorgFit_ABD2232_3 <- mxTryHard(cccAceNorgFit_ABD2232_2)
cccAceNorgFit_ABD2232_4 <- mxTryHard(cccAceNorgFit_ABD2232_3)
cccAceNorgFit_ABD2232_5 <- mxTryHard(cccAceNorgFit_ABD2232_4)
est_cccAceNorgFitABD2232 <- summary(cccAceNorgFit_ABD2232_5)

write.csv(est_cccAceNorgFitABD2232$parameters, "est_cccAceNorgFitABD2232.csv")

# ----THm=THf----

cccAceNothModel_ABD2232 <- mxModel(cccFitABD2232_3, name="cccAceNoth_ABD2232")

cccAceNothModel_ABD2232 <- omxSetParameters( cccAceNothModel_ABD2232,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))

cccAceNothModel_ABD2232 <- omxSetParameters( cccAceNothModel_ABD2232,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi", "t1thmd", "t2thmd", "t1thmq"))

cccAceNothFit_ABD2232 <- mxTryHard(cccAceNothModel_ABD2232)
cccAceNothFit_ABD2232_2 <- mxTryHard(cccAceNothFit_ABD2232)
cccAceNothFit_ABD2232_3 <- mxTryHard(cccAceNothFit_ABD2232_2)

```

```

cccAceNothFit_ABD2232_4 <- mxTryHard(cccAceNothFit_ABD2232_3)

est_cccAceNothFitABD2232 <- summary(cccAceNothFit_ABD2232_2)

write.csv(est_cccAceNothFitABD2232$parameters, "est_cccAceNothFitABD2232.csv")

# ---- equal ACE estimate across sex ----

# equate rg, beta, and ACE across sex

cccAceModel_ABD2232 <- mxModel(cccFitABD2232_3, name="cccAce_ABD2232")

cccAceModel_ABD2232 <- omxSetParameters( cccAceModel_ABD2232, labels=c("rgi", "rgd","rgq"), free=FALSE, values=1 )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("am11","am21","am31", "am22","am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("cm11","cm21","cm31", "cm22","cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("em11","em21","em31", "em22","em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("af11","af21","af31", "af22","af32",
"af33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11","a21","a31", "a22","a32", "a33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("cf11","cf21","cf31", "cf22","cf32",
"cf33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11","c21","c31", "c22","c32", "c33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels=c("ef11","ef21","ef31", "ef22","ef32",
"ef33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11","e21","e31", "e22","e32", "e33" ) )

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceModel_ABD2232 <- omxSetParameters (cccAceModel_ABD2232, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11","b_21","b_31", "b12","b22","b32", "b13","b23","b33"))

cccAceFit_ABD2232 <- mxTryHard(cccAceModel_ABD2232)

cccAceFit_ABD2232_2 <- mxTryHard(cccAceFit_ABD2232)

cccAceFit_ABD2232_3 <- mxTryHard(cccAceFit_ABD2232_2)

cccAceFit_ABD2232_4 <- mxTryHard(cccAceFit_ABD2232_3)

est_cccAceFitABD2232 <- summary(cccAceFit_ABD2232_4)

write.csv(est_cccAceFitABD2232$parameters, "est_cccAceFitABD2232.csv")

# ---- ACE with No beta ----

cccAceNoBModel_ABD2232 <- mxModel(cccAceFit_ABD2232_4, name="cccAceNoB_ABD2232")

cccAceNoBModel_ABD2232 <- omxSetParameters (cccAceNoBModel_ABD2232, labels=c("b_21", "b_31"), free = FALSE,
values = 0)

cccAceNoBFit_ABD2232 <- mxTryHard(cccAceNoBModel_ABD2232)

cccAceNoBFit_ABD2232_2 <- mxTryHard(cccAceNoBFit_ABD2232)

cccAceNoBFit_ABD2232_3 <- mxTryHard(cccAceNoBFit_ABD2232_2)

```

```

est_cccAceNoBFitABD2232 <- summary(cccAceNoBFit_ABD2232_3)

write.csv(est_cccAceNoBFitABD2232$parameters, "est_cccAceNoBFitABD2232.csv")

# ---- ACE drop b_21 ----

cccAceNoB21Model_ABD2232 <- mxModel(cccAceFit_ABD2232_4, name="cccAceNoB21_ABD2232")

cccAceNoB21Model_ABD2232 <- omxSetParameters (cccAceNoB21Model_ABD2232, labels=c("b_21"), free = FALSE,
values = 0)

cccAceNoB21Fit_ABD2232 <- mxTryHard(cccAceNoB21Model_ABD2232)

cccAceNoB21Fit_ABD2232_2 <- mxTryHard(cccAceNoB21Fit_ABD2232)

cccAceNoB21Fit_ABD2232_3 <- mxTryHard(cccAceNoB21Fit_ABD2232_2)

est_cccAceNo21BFitABD2232 <- summary(cccAceNoB21Fit_ABD2232_3)

write.csv(est_cccAceNo21BFitABD2232$parameters, "est_cccAceNo21BFitABD2232.csv")

# ---- ACE drop b_31 ----

cccAceNoB31Model_ABD2232 <- mxModel(cccAceFit_ABD2232_4, name="cccAceNoB31_ABD2232")

cccAceNoB31Model_ABD2232 <- omxSetParameters (cccAceNoB31Model_ABD2232, labels=c("b_31"), free = FALSE,
values = 0)

cccAceNoB31Fit_ABD2232 <- mxTryHard(cccAceNoB31Model_ABD2232)

cccAceNoB31Fit_ABD2232_2 <- mxTryHard(cccAceNoB31Fit_ABD2232)

cccAceNoB31Fit_ABD2232_3 <- mxTryHard(cccAceNoB31Fit_ABD2232_2)

cccAceNoB31Fit_ABD2232_4 <- mxTryHard(cccAceNoB31Fit_ABD2232_3)

cccAceNoB31Fit_ABD2232_5 <- mxTryHard(cccAceNoB31Fit_ABD2232_4)

cccAceNoB31Fit_ABD2232_6 <- mxTryHard(cccAceNoB31Fit_ABD2232_5)

cccAceNoB31Fit_ABD2232_7 <- mxTryHard(cccAceNoB31Fit_ABD2232_6)

est_cccAceNo31BFitABD2232 <- summary(cccAceNoB31Fit_ABD2232_7)

write.csv(est_cccAceNo31BFitABD2232$parameters, "est_cccAceNo31BFitABD2232.csv")

# ---- ACE drop b31, drop C ----

cccAeNoB31Model_ABD2232 <- mxModel(cccAceFit_ABD2232_4, name="cccAeNoB31_ABD2232")

cccAeNoB31Model_ABD2232 <- omxSetParameters (cccAeNoB31Model_ABD2232, labels=c("b_31"), free = FALSE, values
= 0)

cccAeNoB31Model_ABD2232 <- omxSetParameters (cccAeNoB31Model_ABD2232, labels=c("c11", "c22", "c32", "c33"), free
= FALSE, values = 0)

cccAeNoB31Fit_ABD2232 <- mxTryHard(cccAeNoB31Model_ABD2232)

cccAeNoB31Fit_ABD2232_2 <- mxTryHard(cccAeNoB31Fit_ABD2232)

cccAeNoB31Fit_ABD2232_3 <- mxTryHard(cccAeNoB31Fit_ABD2232_2)

cccAeNoB31Fit_ABD2232_4 <- mxTryHard(cccAeNoB31Fit_ABD2232_3)

```

```

est_cccAeNoB31BFitABD2232 <- summary(cccAeNoB31Fit_ABD2232_4)

write.csv(est_cccAeNoB31BFitABD2232$parameters, "est_cccAeNoB31BFitABD2232.csv")

# ---- ACE drop beta 31, drop A ----

cccCeNoB31Model_ABD2232 <- mxModel(cccAceFit_ABD2232_4, name="cccCeNoB31_ABD2232")

cccCeNoB31Model_ABD2232 <- omxSetParameters(cccCeNoB31Model_ABD2232, labels=c("b_31"), free = FALSE, values =
0)

cccCeNoB31Model_ABD2232 <- omxSetParameters(cccCeNoB31Model_ABD2232, labels=c("a11", "a22", "a32", "a33"), free
= FALSE, values = 0)

cccCeNoB31Fit_ABD2232 <- mxTryHard(cccCeNoB31Model_ABD2232)

cccCeNoB31Fit_ABD2232_2 <- mxTryHard(cccCeNoB31Fit_ABD2232)

cccCeNoB31Fit_ABD2232_3 <- mxTryHard(cccCeNoB31Fit_ABD2232_2)

est_cccCeNoB31BFitABD2232 <- summary(cccCeNoB31Fit_ABD2232_3)

write.csv(est_cccCeNoB31BFitABD2232$parameters, "est_cccCeNoB31BFitABD2232.csv")

# ---- ACE drop beta 31, drop C33 ----

cccACeNoB31A33Model_ABD2232 <- mxModel(cccAceFit_ADH1619_4, name="cccACeNoB31A33_ABD2232")

cccACeNoB31A33Model_ABD2232 <- omxSetParameters(cccACeNoB31A33Model_ABD2232, labels=c("b_31"), free =
FALSE, values = 0)

cccACeNoB31A33Model_ABD2232 <- omxSetParameters(cccACeNoB31A33Model_ABD2232, labels=c("c33"), free =
FALSE, values = 0)

cccACeNoB31A33Fit_ABD2232 <- mxTryHard(cccACeNoB31A33Model_ABD2232)

cccACeNoB31A33Fit_ABD2232_2 <- mxTryHard(cccACeNoB31A33Fit_ABD2232)

est_cccACeNoB31A33Fit_ABD2232_2 <- summary(cccACeNoB31A33Fit_ABD2232_2)

write.csv(est_cccACeNoB31A33Fit_ABD2232_2$parameters, "est_cccACeNoB31A33Fit_ABD2232_2.csv")

# ---- All submodels ----

fitstatsubmodels1_ABD2232 <- mxCompare(cccFitABD2232_3, nested <- list(cccAceNorgFit_ABD2232_5,
cccAceNothFit_ABD2232_2,
cccAceFit_ABD2232_4))

fitstatsubmodels2_ABD2232 <- mxCompare(cccAceFit_ABD2232_4, nested <- list(cccAceNoBFit_ABD2232_3,
cccAceNoB21Fit_ABD2232_3,
cccAceNoB31Fit_ABD2232_7,
cccAeNoB31Fit_ABD2232_4,
cccCeNoB31Fit_ABD2232_3))

fitstatsubmodels_ABD2232 <- rbind(fitstatsubmodels1_ABD2232, fitstatsubmodels2_ABD2232)

write.csv(fitstatsubmodels_ABD2232, "submodelsABD2232.csv")

```

```

# base model

write.csv(summary(cccFitABD2232_3)$parameters, "estimates_basemodel_ABD_2232.csv")

# ----- Create Confidence Interval Objects: best fitting model -----

# run the best fitting with 95% CIs

bestci <- mxCI( c("a11", "a22", "a32", "a33",
                "c11", "c22", "c32", "c33",
                "e11", "e22", "e32", "e33",
                "b_21"))

bestmodel_ABD_2232 <- mxModel(cccAceNoB31Fit_ABD2232_7, bestci)

fitbest_ABD_2232 <- mxTryHard(bestmodel_ABD_2232, intervals = F)
fitbest_ABD_2232_2 <- mxTryHard(fitbest_ABD_2232, intervals = F)
fitbest_ABD_2232_3 <- mxTryHard(fitbest_ABD_2232_2, intervals = F)
fitbest_ABD_2232_int <- mxRun(fitbest_ABD_2232_3, intervals = T)
sumbest_ABD_2232 <- summary(fitbest_ABD_2232_int)

sumbest_ABD_2232$parameters

sumbest_ABD_2232$CI

sumbest_ABD_2232$CI$CIint <- paste("(",round(sumbest_ABD_2232$CI$lbound,2), "-",
round(sumbest_ABD_2232$CI$ubound,2),")", sep="")

# CI of best fitting model

write.csv(sumbest_ABD_2232$CI, "CI_bestfitting_ABD_2232.csv")

# ----- Create Confidence Interval Objects: ACE model no sex dif-----

ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                "c11", "c22", "c32", "c33",
                "e11", "e22", "e32", "e33",
                "b_21", "b_31"))

ACEmodel_ABD_2232 <- mxModel(cccAceFit_ABD2232_4, ACEciACEnorg)

fitACE_ABD_2232 <- mxTryHard(ACEmodel_ABD_2232, intervals = F)
fitACE_ABD_2232_2 <- mxTryHard(fitACE_ABD_2232, intervals = F)
fitACE_ABD_2232_3 <- mxTryHard(fitACE_ABD_2232_2, intervals = F)
fitACE_ABD_2232_int <- mxRun(fitACE_ABD_2232_3, intervals = T)
sumACE_ABD_2232 <- summary(fitACE_ABD_2232_int)

sumACE_ABD_2232$parameters

```

```
sumACE_ABD_2232$CI
sumACE_ABD_2232$CI$CIint <- paste("(",round(sumACE_ABD_2232$CI$lbound,2), "-",
round(sumACE_ABD_2232$CI$subound,2),")", sep="")
sumACE_ABD_22321 <- sumACE_ABD_2232$CI
sumACE_ABD_22321$datagroup <- "ABD_2232"
write.csv(sumACE_ABD_22321, "sumACE_ABD_2232.csv")
```


Appendix X R script for CCC Model for VTSABD/TSA all participants. (ABD: VTSABD/TSA)

```
# -----Load Library-----
require(OpenMx)
require(psych)
source("http://www.vipbg.vcu.edu/~vipbg/GE/GenEpiHelperFunctions.R")
library('polycor')
mxOption( NULL, "Default optimizer", "NPSOL" )
# -----
# ----- PREPARE DATA -----
data <- read.table("ABDbuild.csv", header=TRUE, sep=",")
twindat <- function(dat, famid, twinid, zygosity) {
  datA <- dat[dat[,twinid]==min(dat[,twinid]),] #twin1
  datB <- dat[dat[,twinid]==max(dat[,twinid]),] #twin2
  DAT <- merge(datA, datB, by=famid, all.x=TRUE, all.y=TRUE, suffixes=c("_T1", "_T2"))
  DAT[,paste(twinid, "_T1", sep="")] <- NULL
  DAT[,paste(twinid, "_T2", sep="")] <- NULL
  DAT[,zygosity] <-
  ifelse(is.na(DAT[,paste(zygosity, "_T1", sep="")]), DAT[,paste(zygosity, "_T2", sep="")], DAT[,paste(zygosity, "_T1", sep="")])
  DAT[,paste(zygosity, "_T1", sep="")] <- NULL
  DAT[,paste(zygosity, "_T2", sep="")] <- NULL
  return(DAT)
}
# three arguments needed for twindat() function
nic <- twindat(dat=data, famid= "FAMNO", twinid= "IndId_int", zygosity= "zygroup")
# Load Data
describe(nic, skew=F)
# Select Variables for Analysis
Vars <- c('tobever', 'maxcigxdayCCC', 'quiteverCCC')
nv <- 3 # number of variables
ntv <- nv*2 # number of total variables
selVars <- paste(Vars, c(rep("_T1", nv), rep("_T2", nv)), sep="")
```

```

# Specify Thresholds for Ordinal Variables

## nth: number of thresholds; fcat: first category; lcat: last category; ncat: number of categories;

nth1  <- 1 ; fcat1  <- 0 ; lcat1  <- fcat1+nth1 ; ncat1  <- nth1+1
nth2  <- 2 ; fcat2  <- 0 ; lcat2  <- fcat2+nth2 ; ncat2  <- nth2+1
nth3  <- 1 ; fcat3  <- 0 ; lcat3  <- fcat3+nth3 ; ncat3  <- nth3+1
nth   <- max(nth1,nth2,nth3)

# Specify Arguments for Threshold Matrices

## lth: lowest threshold; ith: increment;

lth1  <- 0 ; ith1  <- 0.1 ;
lth2  <- 0 ; ith2  <- 0.1 ;
lth3  <- 0 ; ith3  <- 0.1 ;

thFree <- c(rep(T,nth1),rep(F,nth-nth1),rep(T,nth2),rep(F,nth-nth2),rep(T,nth3),rep(F,nth-nth3))
thValues <- matrix(c(lth1,(rep(ith1,nth-1)),lth2,(rep(ith2,nth-1)),lth3,(rep(ith3,nth-1))),nrow=nth,ncol=nv)
thLBound <- matrix(rep(c(-3,(rep(0.001,nth-1))),nv),nrow=nth,ncol=nv)

# Select Data for Analysis

twinData <- nic[,c(selVars,'zygroup')]
describe(twinData)

twinDataBin <- twinData

# Factorize Ordinal Variables

twinDataBin[,c(1,nv+1)] <- mxFactor(twinDataBin[,c(1,nv+1)], levels = c(0:nth1))
twinDataBin[,c(2,nv+2)] <- mxFactor(twinDataBin[,c(2,nv+2)], levels = c(0:nth2))
twinDataBin[,c(3,nv+3)] <- mxFactor(twinDataBin[,c(3,nv+3)], levels = c(0:nth3))

# Create Datasets by Zygosity- 5 group

dataBinMZm <- subset(twinDataBin, zygroup==1, selVars)
dataBinMZf <- subset(twinDataBin, zygroup==2, selVars)
dataBinDZm <- subset(twinDataBin, zygroup==3, selVars)
dataBinDZf <- subset(twinDataBin, zygroup==4, selVars)
dataBinDZo <- subset(twinDataBin, zygroup==5, selVars)

# ---- Prepare Genetic Model ----
# -----
# PREPARE GENETIC MODEL
# -----

```

```

# Heterogeneity CCC ACE Model

# -----

# Specify Arguments for Causal Path Matrices

svB    <- 0

freeBs <- c(F,T,T, F,F,F, F,F,F)

svBs   <- c(0,svB, svB, 0, 0, 0, 0, 0, 0)

labBsm <- c("b11","bm21","bm31", "b12","b22","b32", "b13","b23","b33")

labBsf <- c("b11","bf21","bf31", "b12","b22","b32", "b13","b23","b33")

# Matrices a, c, and e to store a, c, and e path coefficients

pathAm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("am11","am21","am31","am22","am32","am33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="am" )

pathCm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cm11","cm21","cm31","cm22","cm32","cm33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cm" )

pathEm <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("em11","em21","em31","em22","em32","em33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="em" )

pathAf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("af11","af21","af31","af22","af32","af33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="af" )

pathCf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("cf11","cf21","cf31","cf22","cf32","cf33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="cf" )

pathEf <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(T,F,F, T,T, T), values=c(0.6,0,0, 0.6, 0.6, 0.6),
labels=c("ef11","ef21","ef31","ef22","ef32","ef33"), lbound=c(-1,0, 0.0001,-1,0.0001, 0.0001), ubound=1, name="ef" )

pathRg <- mxMatrix( type="Diag", nrow=nv, ncol=nv, free=c(T,T,T), values=1, labels=c("rgi","rgd","rgq"), lbound=-1,
ubound=1, name="rg" )

pathBm <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsm, lbound=-.99, ubound=.99,
name="bm" )

pathBf <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=freeBs, values=svBs, labels=labBsf, lbound=-.99, ubound=.99,
name="bf" )

# Matrices A, C, and E compute variance components

covAm <- mxAlgebra( expression=am %*% t(am), name="Am" )

covCm <- mxAlgebra( expression=cm %*% t(cm), name="Cm" )

covEm <- mxAlgebra( expression=em %*% t(em), name="Em" )

covAf <- mxAlgebra( expression=af %*% t(af), name="Af" )

covCf <- mxAlgebra( expression=cf %*% t(cf), name="Cf" )

covEf <- mxAlgebra( expression=ef %*% t(ef), name="Ef" )

# Algebra to compute total variances and standard deviations (diagonal only)

nvl   <- mxMatrix( type="Iden", nrow=nv, ncol=nv, name="nvl" )

```

```

nvZ    <- mxMatrix( type="Zero", nrow=nv, ncol=nv, name="nvZ" )
I2     <- mxMatrix( type="Iden", nrow=2, ncol=2, name="I2" )
Unv1   <- mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
Vm     <- mxAlgebra( expression=(solve(nvI-bm)) %&% (Am+Cm+Em), name="Vm" )
Vf     <- mxAlgebra( expression=(solve(nvI-bf)) %&% (Af+Cf+Ef), name="Vf" )
iSDm   <- mxAlgebra( expression=solve(sqrt(nvI*Vm)), name="iSDm" )
iSDf   <- mxAlgebra( expression=solve(sqrt(nvI*Vf)), name="iSDf" )

# Constraint on variance of Binary variables

Var1m  <- mxConstraint( expression=diag2vec(Vm)==Unv1, name="Var1m" )
Var1f  <- mxConstraint( expression=diag2vec(Vf)==Unv1, name="Var1f" )

# Matrix & Algebra for expected means vector and expected thresholds

Mean   <- mxMatrix( type="Zero", nrow=1, ncol=nv, name="Mean" )
expMean <- mxAlgebra( expression= cbind(Mean,Mean), name="expMean" )

Inc    <- mxMatrix( type="Lower", nrow=nth, ncol=nth, free=FALSE, values=1, name="Inc" )

Threm  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmim",sep=""),paste("t",1:nth,"thmdm",sep=""),paste("t",1:nth,"thmqm",sep="")), lbound=thLBound,
name="Threm" )

Thref  <- mxMatrix( type="Full", nrow=nth, ncol=nv, free=thFree, values=thValues,
labels=c(paste("t",1:nth,"thmif",sep=""),paste("t",1:nth,"thmdf",sep=""),paste("t",1:nth,"thmqf",sep="")), lbound=thLBound,
name="Thref" )

ThIncm <- mxAlgebra( expression= Inc %*% Threm, name="ThIncm" )
ThIncf <- mxAlgebra( expression= Inc %*% Thref, name="ThIncf" )

expThreZm <- mxAlgebra( expression= cbind(ThIncm,ThIncm), name="expThreZm" )
expThreZf <- mxAlgebra( expression= cbind(ThIncf,ThIncf), name="expThreZf" )
expThreZo <- mxAlgebra( expression= cbind(ThIncm,ThIncf), name="expThreZo" )

# Algebra for expected variance/covariance matrix

expCovMZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , Am+Cm),
cbind(Am+Cm , Am+Cm+Em)), name="expCovMZm" )

expCovMZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , Af+Cf), cbind(Af+Cf ,
Af+Cf+Ef)), name="expCovMZf" )

expCovDZm <- mxAlgebra( expression=( I2 %x% solve(nvI-bm) ) %&% rbind( cbind(Am+Cm+Em , 0.5%x%Am+Cm),
cbind(0.5%x%Am+Cm , Am+Cm+Em)), name="expCovDZm" )

expCovDZf <- mxAlgebra( expression=( I2 %x% solve(nvI-bf) ) %&% rbind( cbind(Af+Cf+Ef , 0.5%x%Af+Cf),
cbind(0.5%x%Af+Cf , Af+Cf+Ef)), name="expCovDZf" )

expCovDZo <- mxAlgebra( expression=( rbind( cbind(solve(nvI-bm), nvZ), cbind(nvZ, solve(nvI-bf))) ) %&%
rbind( cbind(Am+Cm+Em , 0.5%x%(rg*(am%*%t(af))+cm%*%t(cf)),

```

```

cbind(0.5%x%(rg*(af%*%(am)))+cf%*%(cm) , Af+Cf+Ef), name="expCovDZo" )

# Data objects for Multiple Groups
dataMZf <- mxData( observed=dataBinMZf, type="raw" )
dataDZf <- mxData( observed=dataBinDZf, type="raw" )
dataMZm <- mxData( observed=dataBinMZm, type="raw" )
dataDZm <- mxData( observed=dataBinDZm, type="raw" )
dataDZo <- mxData( observed=dataBinDZo, type="raw" )

# Objective objects for Multiple Groups
expMZf <- mxExpectationNormal( covariance="expCovMZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expMZm <- mxExpectationNormal( covariance="expCovMZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZf <- mxExpectationNormal( covariance="expCovDZf", means="expMean", dimnames=selVars,
thresholds="expThreZf")
expDZm <- mxExpectationNormal( covariance="expCovDZm", means="expMean", dimnames=selVars,
thresholds="expThreZm")
expDZo <- mxExpectationNormal( covariance="expCovDZo", means="expMean", dimnames=selVars,
thresholds="expThreZo")
funML <- mxFitFunctionML()

# Combine Groups
pars <- list( nvI, nvZ, I2, Unv1, Mean, expMean, Inc )
parsZf <- list( pathAf, pathCf, pathEf, pathBf, covAf, covCf, covEf, Vf, iSDf, Thref, ThIncf)
parsZm <- list( pathAm, pathCm, pathEm, pathBm, covAm, covCm, covEm, Vm, iSDm, Threm, ThIncm)
modelMZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovMZf, dataMZf, expMZf, funML, name="MZf" )
modelDZf <- mxModel( pars, parsZf, expMean, expThreZf, expCovDZf, dataDZf, expDZf, funML, name="DZf" )
modelMZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovMZm, dataMZm, expMZm, funML, name="MZm" )
modelDZm <- mxModel( pars, parsZm, expMean, expThreZm, expCovDZm, dataDZm, expDZm, funML, name="DZm" )
modelDZo <- mxModel( pars, parsZf, pathRg, parsZm, expMean, expThreZo, expCovDZo, dataDZo, expDZo, funML,
name="DZo" )
multi <- mxFitFunctionMultigroup(c('MZf', 'MZm', 'DZf', 'DZm', 'DZo' ))
cccModelABD_all <- mxModel( "ccc_ABD_all", pars, parsZf, parsZm, Var1m, Var1f, modelMZf, modelDZf, modelMZm,
modelDZm, modelDZo, multi)

# -----
# ----- RUN GENETIC MODEL -----

```

```

# Run CCC and Cholesky Decomposition ACE model

cccFitABD_all <- mxTryHard(cccModelABD_all, intervals=F)
cccFitABD_all_2 <- mxTryHard(cccFitABD_all, intervals=F)
cccFitABD_all_3 <- mxTryHard(cccFitABD_all_2, intervals=F)
cccFitABD_all_4 <- mxTryHard(cccFitABD_all_3, intervals=F)
est_cccFitABD_all <- summary(cccFitABD_all_2)
write.csv(est_cccFitABD_all$parameters, "est_cccFitABD_all.csv")

# Test Submodels constraining sex

# -----
# ---- No rg ----

cccAceNorgModel_ABD_all <- mxModel(cccFitABD_all_2, name="cccAceNorg_ABD_all")

cccAceNorgModel_ABD_all <- omxSetParameters( cccAceNorgModel_ABD_all, labels=c("rgi","rgd","rgq"), free=FALSE,
values=1 )

cccAceNorgFit_ABD_all <- mxTryHard(cccAceNorgModel_ABD_all)
cccAceNorgFit_ABD_all_2 <- mxTryHard(cccAceNorgFit_ABD_all)
cccAceNorgFit_ABD_all_3 <- mxTryHard(cccAceNorgFit_ABD_all_2)
cccAceNorgFit_ABD_all_4 <- mxTryHard(cccAceNorgFit_ABD_all_3)
cccAceNorgFit_ABD_all_5 <- mxTryHard(cccAceNorgFit_ABD_all_4)
cccAceNorgFit_ABD_all_6 <- mxTryHard(cccAceNorgFit_ABD_all_5)
est_cccAceNorgFitABD_all <- summary(cccAceNorgFit_ABD_all_5)
write.csv(est_cccAceNorgFitABD_all$parameters, "est_cccAceNorgFitABD_all.csv")

# ----THm=THf----

cccAceNothModel_ABD_all <- mxModel(cccFitABD_all_2, name="cccAceNoth_ABD_all")
cccAceNothModel_ABD_all <- omxSetParameters( cccAceNothModel_ABD_all,
labels=c("t1thmim", "t1thmdm", "t1thmqm","t2thmdm"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))
cccAceNothModel_ABD_all <- omxSetParameters( cccAceNothModel_ABD_all,
labels=c("t1thmif", "t1thmdf", "t1thmqf","t2thmdf"), free=thFree, values=thValues,
newlabels=c("t1thmi","t1thmd", "t2thmd","t1thmq"))

cccAceNothFit_ABD_all <- mxTryHard(cccAceNothModel_ABD_all)
cccAceNothFit_ABD_all_2 <- mxTryHard(cccAceNothFit_ABD_all)
cccAceNothFit_ABD_all_3 <- mxTryHard(cccAceNothFit_ABD_all_2)

```

```

cccAceNothFit_ABD_all_4 <- mxTryHard(cccAceNothFit_ABD_all_3)
cccAceNothFit_ABD_all_5 <- mxTryHard(cccAceNothFit_ABD_all_4)
est_cccAceNothFitABD_all <- summary(cccAceNothFit_ABD_all_5)
write.csv(est_cccAceNothFitABD_all$parameters, "est_cccAceNothFitABD_all.csv")

# ---- equal ACE estimate across sex ----
# equate rgq, beta, and ACE across sex

cccAceModel_ABD_all <- mxModel(cccFitABD_all_2, name="cccAce_ABD_all")
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("rgi", "rgd", "rgq"), free=FALSE, values=1 )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("am11", "am21", "am31", "am22", "am32",
"am33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("cm11", "cm21", "cm31", "cm22", "cm32",
"cm33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("em11", "em21", "em31", "em22", "em32",
"em33" ), free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("af11", "af21", "af31", "af22", "af32", "af33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("a11", "a21", "a31", "a22", "a32", "a33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("cf11", "cf21", "cf31", "cf22", "cf32", "cf33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("c11", "c21", "c31", "c22", "c32", "c33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels=c("ef11", "ef21", "ef31", "ef22", "ef32", "ef33" ),
free=c(T,F,F, T,T, T), values=c(0.6, 0, 0, 0.6, 0.6, 0.6), newlabels=c("e11", "e21", "e31", "e22", "e32", "e33" ) )
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels = labBsm, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))
cccAceModel_ABD_all <- omxSetParameters( cccAceModel_ABD_all, labels = labBsf, free = freeBs, values = svBs,
newlabels=c("b11", "b_21", "b_31", "b12", "b22", "b32", "b13", "b23", "b33"))
cccAceFit_ABD_all <- mxTryHard(cccAceModel_ABD_all)
cccAceFit_ABD_all_2 <- mxTryHard(cccAceFit_ABD_all)
cccAceFit_ABD_all_3 <- mxTryHard(cccAceFit_ABD_all_2)
est_cccAceFitABD_all <- summary(cccAceFit_ABD_all_3)
write.csv(est_cccAceFitABD_all$parameters, "est_cccAceFitABD_all.csv")

# ---- ACE with No beta ----
cccAceNoBModel_ABD_all <- mxModel(cccAceFit_ABD_all_3, name="cccAceNoB_ABD_all")
cccAceNoBModel_ABD_all <- omxSetParameters( cccAceNoBModel_ABD_all, labels=c("b_21", "b_31"), free = FALSE,
values = 0)
cccAceNoBFit_ABD_all <- mxTryHard(cccAceNoBModel_ABD_all)
cccAceNoBFit_ABD_all_2 <- mxTryHard(cccAceNoBFit_ABD_all)
cccAceNoBFit_ABD_all_3 <- mxTryHard(cccAceNoBFit_ABD_all_2)

```

```

est_cccAceNoBFitABD_all <- summary(cccAceNoBFit_ABD_all_3)

write.csv(est_cccAceNoBFitABD_all$parameters, "est_cccAceNoBFitABD_all.csv")

# ---- ACE drop b_21 ----

cccAceNoB21Model_ABD_all <- mxModel(cccAceFit_ABD_all_3, name="cccAceNoB21_ABD_all")

cccAceNoB21Model_ABD_all <- omxSetParameters (cccAceNoB21Model_ABD_all, labels=c("b_21"), free = FALSE, values =
0)

cccAceNoB21Fit_ABD_all <- mxTryHard(cccAceNoB21Model_ABD_all)

cccAceNoB21Fit_ABD_all_2 <- mxTryHard(cccAceNoB21Fit_ABD_all)

cccAceNoB21Fit_ABD_all_3 <- mxTryHard(cccAceNoB21Fit_ABD_all_2)

est_cccAceNo21BFitABD_all <- summary(cccAceNoB21Fit_ABD_all_3)

write.csv(est_cccAceNo21BFitABD_all$parameters, "est_cccAceNo21BFitABD_all.csv")

# ---- ACE drop b_31 ----

cccAceNoB31Model_ABD_all <- mxModel(cccAceFit_ABD_all_3, name="cccAceNoB31_ABD_all")

cccAceNoB31Model_ABD_all <- omxSetParameters (cccAceNoB31Model_ABD_all, labels=c("b_31"), free = FALSE, values =
0)

cccAceNoB31Fit_ABD_all <- mxTryHard(cccAceNoB31Model_ABD_all)

cccAceNoB31Fit_ABD_all_2 <- mxTryHard(cccAceNoB31Fit_ABD_all)

cccAceNoB31Fit_ABD_all_3 <- mxTryHard(cccAceNoB31Fit_ABD_all_2)

cccAceNoB31Fit_ABD_all_4 <- mxTryHard(cccAceNoB31Fit_ABD_all_3)

est_cccAceNo31BFitABD_all <- summary(cccAceNoB31Fit_ABD_all)

write.csv(est_cccAceNo31BFitABD_all$parameters, "est_cccAceNo31BFitABD_all.csv")

# ---- ACE, drop C: WITH ALL B ----

cccAeNoB31Model_ABD_all <- mxModel(cccAceFit_ABD_all_3, name="cccAe_ABD_all")

cccAeNoB31Model_ABD_all <- omxSetParameters (cccAeNoB31Model_ABD_all, labels=c("c11", "c22", "c32", "c33"), free =
FALSE, values = 0)

cccAeNoB31Fit_ABD_all <- mxTryHard(cccAeNoB31Model_ABD_all)

cccAeNoB31Fit_ABD_all_2 <- mxTryHard(cccAeNoB31Fit_ABD_all)

cccAeNoB31Fit_ABD_all_3 <- mxTryHard(cccAeNoB31Fit_ABD_all_2)

est_cccAeNo31BFitABD_all <- summary(cccAeNoB31Fit_ABD_all_3)

write.csv(est_cccAeNo31BFitABD_all$parameters, "est_cccAeNo31BFitABD_all.csv")

# ---- ACE drop A: WITH ALL B ----

cccCeNoB31Model_ABD_all <- mxModel(cccAceFit_ABD_all_3, name="cccCe_ABD_all")

cccCeNoB31Model_ABD_all <- omxSetParameters (cccCeNoB31Model_ABD_all, labels=c("a11", "a22", "a32", "a33"), free =
FALSE, values = 0)

```



```

cccCeNoB31Fit_ABD_all <- mxTryHard(cccCeNoB31Model_ABD_all)
cccCeNoB31Fit_ABD_all_2 <- mxTryHard(cccCeNoB31Fit_ABD_all)
cccCeNoB31Fit_ABD_all_3 <- mxTryHard(cccCeNoB31Fit_ABD_all_2)
est_cccCeNo31BFitABD_all <- summary(cccCeNoB31Fit_ABD_all_3)
write.csv(est_cccCeNo31BFitABD_all$parameters, "est_cccCeNo31BFitABD_all.csv")
# ---- All submodels ----
fitstatsubmodels1_ABD_all <- mxCompare(cccFitABD_all_2, nested <- list(cccAceNorgFit_ABD_all_5,
                                                                    cccAceNothFit_ABD_all_5,
                                                                    cccAceFit_ABD_all_3))
fitstatsubmodels2_ABD_all <- mxCompare(cccAceFit_ABD_all_3, nested <- list(cccAceNoBFit_ABD_all_3,
                                                                    cccAceNoB21Fit_ABD_all_3,
                                                                    cccAceNoB31Fit_ABD_all,
                                                                    cccAeNoB31Fit_ABD_all_3,
                                                                    cccCeNoB31Fit_ABD_all_3))
fitstatsubmodels_ABD_all <- rbind(fitstatsubmodels1_ABD_all, fitstatsubmodels2_ABD_all)
write.csv(fitstatsubmodels_ABD_all, "submodelsABD_all.csv")
# base model
write.csv(summary(cccFitABD_all_2)$parameters, "estimates_basemodel_ABD_all.csv")
# ----- Create Confidence Interval Objects: best fitting and full model -----
ACEciACEnorg <- mxCI( c("a11", "a22", "a32", "a33",
                       "c11", "c22", "c32", "c33",
                       "e11", "e22", "e32", "e33",
                       "b_21", "b_31"))
ACEmodel_ABD_all <- mxModel(cccAceFit_ABD_all_3, ACEciACEnorg)
fitACE_ABD_all <- mxTryHard(ACEmodel_ABD_all, intervals = F)
fitACE_ABD_all_2 <- mxTryHard(fitACE_ABD_all, intervals = F)
fitACE_ABD_all_3 <- mxTryHard(fitACE_ABD_all_2, intervals = F)
fitACE_ABD_all_int <- mxRun(fitACE_ABD_all_3, intervals = T)
sumACE_ABD_all <- summary(fitACE_ABD_all_int)
sumACE_ABD_all$parameters
sumACE_ABD_all$CI

```

```
sumACE_ABD_all$CI$CIint <- paste("(",round(sumACE_ABD_all$CI$lbound,2), "-",  
round(sumACE_ABD_all$CI$subound,2),")", sep="")  
  
sumACE_ABD_all1 <- sumACE_ABD_all$CI  
  
sumACE_ABD_all1$datagroup <- "ABD_all"  
  
write.csv(sumACE_ABD_all1, "sumbestandACE_ABD_all.csv")
```

Vita

Gladys Emmanuella Putri Langi was born on September 27, 1989, in Bandung, West Java, Indonesia and is an Indonesian citizen. She graduated from Sekolah Menengah Atas Kristen Badan Pendidikan Kristen Penabur Bandung (High School), Indonesia, in 2008. She received her Bachelor of Science in Microbiology, with a thesis on molecular genetics, from Institut Teknologi Bandung, Indonesia in 2012. She then worked at Mochtar Riady Institute for Nanotechnology as a research assistant studying human genetic variations for three years.