THE INTERPLAY OF GENES, PEERS, AND COHORT ON PATTERNS OF CIGARETTE
USE

by

AMANDA G. WILLS

B.S., University of Texas at Austin, 2009

M.A., University of Colorado at Boulder, 2013

April 1, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2015

This thesis entitled:
The Interplay of Genes, Peers, and Cohort on Patterns of Cigarette Use
written by Amanda Golzar Wills
has been approved for the Department of Psychology and Neuroscience

_____
Gregory Carey

_____
Matthew C. Keller

_____
Matthew B. McQueen

_____
Michael Stallings

_____
Angela Bryan

*Date*_____

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

Wills, Amanda Golzar (Ph.D., Department of Psychology and Neuroscience)

The Interplay of Genes, Peers, and Cohort on Patterns of Cigarette Use

Thesis directed by Associate Professor Gregory Carey and Associate Professor Matthew C. Keller

Genetic factors are a known culprit influencing the choice of individuals to smoke. However, the way in which these genetic factors may contribute to smoking behavior through peers, depending upon the specific stage of smoking, and change in relation to birth cohort further compounds the complexity of understanding the mechanisms that link genetic makeup to smoking behavior.

Here we addressed: (1) what mechanism(s) are responsible for peer similarity in smoking behavior, (2) at what stage of the smoking trajectory, from initiation to progression, are these peer mechanisms most salient, (3) from whole-genome SNP data, to what extent is smoking initiation related to more regular smoking behaviors, and (4) to what degree do the genetic factors influencing smoking for one generation correspond to those of another.

We utilized two twin samples, the 1962 National Merit twins and the more recent Add Health twins. We also conducted genome-wide analyses of data from the Atherosclerosis Risk in Communities Study (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA).

Our results indicated that homophily, or the tendency to associate with individuals that are like oneself, may explain peer homogeneity for smoking

behaviors, and if this homophily is accompanied by additional peer influence, active gene-environment correlation may be in part responsible for peer resemblance in smoking behavior. While it was unclear whether this mechanism is relevant to both initiation and persistent smoking in the National Merit Twins, analysis on the Add Health sample demonstrated that this mechanism may be important at both the stages of experimentation and regular use.

Genome-wide analysis on unrelated individuals revealed that common genetic variation, as indexed by genome-wide SNPs, contributed to cigarette smoking liability, and the genetic factors that influenced smoking initiation were largely shared with those that impacted quantity smoked. Additionally, the genetic factors influencing smoking may change as a function of birth cohort.

ACKNOWLEDGEMENTS

I would also like to thank my husband Mike, my parents, and my friends who have provided so much love and support throughout this process.

CONTENTS

TABLES

Table

CHAPTER I

INTRODUCTION

The most recent Surgeon General's report reiterated findings that cigarette

smoking has been linked to a wide range of health deficits including diabetes,

arthritis, cardiovascular disease, and various cancers (National Center for Chronic

Disease Prevention and Health Promotion (US) Office on Smoking and Health 2014).

Since the first Surgeon General's report was published in 1964, 20 million

Americans have died as a result of smoking, deeming it one of the 'greatest public

health catastrophes of the century.' While this report noted that rates of smoking

have declined, it is still a public health problem, especially for certain groups of

individuals. Thus, even as the evidence for the numerous health consequences of

cigarette smoking builds, individuals continue to choose to smoke.

Understanding the mechanisms behind choices to undertake in actions

detrimental to health is one key to developing ways to effectively prevent these

choices. Thus, researchers have attempted to elucidate the various mechanisms that

contribute to cigarette smoking behavior. Some known culprits include

sociocultural norms, peer relationships, and genetic influences (Kobus 2003;

Munafò & Johnstone, 2008; Van Reek & Drop, 1986). However, these mechanisms

are unlikely to work in isolation; further, the influences related to starting to smoke

may be different from those related to more heavy patterns of cigarette smoking.

Our primary focus is the interface between genetic factors, peer influences, and

cohort effects, and how these factors may differ depending on how the smoking phenotype is defined.

Literature Review

Peers and Cigarette Smoking

Association with cigarette smoking peers has been shown to increase the likelihood that one will also smoke (Alexander, Piazza, Mekos, & Valente, 2001; Hoffman, Sussman, Unger, & Valente, 2006; Holliday, Rothwell, & Moore, 2010; Pollard, Tucker, Green, Kennedy, & Go, 2010; Vink, Willemsen, Engels, & Boomsma, 2003). For example, belonging to peer networks composed of at least half smokers or having best friends who smoked is associated with individuals being twice as likely to smoke (Alexander et al., 2001). This observed homogeneity in smoking behavior between individuals and their peer groups was reviewed by Kobus (2003) who summarized the work done to pinpoint the specifics of *how* peers are related to smoking behavior, but concluded that further work should be done to illuminate the 'subtleties' behind the social dynamics of this contribution.

Therefore, we know that individuals tend to resemble their peers in smoking behavior, yet we know little about why this association of smoking behavior, or homogeneity, between adolescents and their peer groups is observed. Two potential mechanisms may influence the correlation in smoking among peers—*homophily* and peer influence. Homophily is defined as the propensity to associate with more similar than dissimilar individuals (McPherson, Smith-Lovin, & Cook, 2001). McPherson et al. noted that it was originally assumed that peer groups directly influenced an individual's behavior. However, the surge of longitudinal data led to a

shift towards recognition of the importance of homophily: individuals may actually select membership into groups that share one's initial behavioral characteristics. Thus, in the context of homogeneity of smoking between individuals and peer groups, arise two distinct, yet non-mutually exclusive, possibilities: (1) peer groups may directly influence smoking behavior (peer influence), and (2) those with a propensity for smoking behavior may self-select into groups with similar characteristics (homophily).

However, there remains debate over the relative contribution of each of these possibilities. For example, Arnett (2007) rejected the assumption that the association between peer and individual smoking is a result of direct peer influence, and suggested that selection of friends based on a number of factors leads to peer group selection that creates a pathway to peer context variables such as group expectations, identity, and opportunities that may influence smoking behavior. Simons-Morton & Farhat (2010) on the other hand conclude that both homophily and peer influence are important. Most reports emphasize that the magnitude of each influence remains unclear (Dishion & Owen, 2002; Go, Green, Kennedy, Pollard, & Tucker, 2010; Hall & Valente, 2007; Mercken, Snijders, Steglich, Vertiainen, & de Vries, 2010; White, Hopper, Wearing, & Hill, 2003).

Further, the effect of peers on smoking behavior may vary as a function of how smoking is defined, as differing etiologies have been implicated for smoking initiation versus regular smoking and nicotine dependence (Amos, Spitz, & Cinciripini, 2010; Vink, Willemsen, & Boomsma, 2005). Thus, we attempt to gain a

better understanding of the peer contribution to smoking behavior and how this contribution may differ depending on the specific smoking phenotype.

The Genetics of Cigarette Smoking

Genetic factors have also been shown to play a role in cigarette smoking. Such evidence for a genetic basis to smoking behavior can be traced back to the late 1950s and 1960s (Hughes 1986). A number of these studies reported that the smoking behaviors of monozygotic (MZ) twins were more concordant compared to dizygotic (DZ) twins (Fisher 1958; Friberg, Kaij, Dencker, & Jonsson, 1959; Todd & Mason, 1959). More recent analyses have attempted to understand and quantify the genetic basis of smoking even further. Sullivan and Kendler (1999) reviewed the existing literature that included twin, family, and adoption studies, and demonstrated that the etiology of smoking behavior may depend on whether smoking is defined as experimentation, initiation or progression to nicotine dependence. The major finding from this review was that genetic factors contributed substantially to both smoking initiation (mean heritability = .56) and nicotine dependence (mean heritability = .67). While unique environment (and measurement error) was important for both initiation and nicotine dependence, shared environment played a role in smoking initiation but was negligible for nicotine dependence.

Whole genome methods. Given the overwhelming evidence for a genetic contribution to smoking using twin samples, attempts to pinpoint specific genetic regions related to smoking using genome-wide association analysis (GWAS) have been puzzling in their failure to account for any sizable proportion of the heritability

reported in twin studies. This problem of the 'missing heritability' is not unique to the smoking phenotype, and it has been suggested that one way to potentially improve future GWAS analyses would be to increase sample size and focus on meta-analyses (Maher 2008; Manolio et al., 2009). Regarding smoking phenotypes, the Tobacco and Genetics Consortium (2010) combined three consortia to amass a sample size of over 140,000 individuals. While their findings did pinpoint biologically relevant markers, the loci identified were still unable to account for even a small proportion of previously estimated heritability.

In contrast to genome-wide analyses, that attempt to pinpoint specific regions or SNPs that contribute to trait variance, Yang et al. (2010) developed a method known as Genome-wide Complex Trait Analysis (GCTA) that estimates the trait variance explained by all genome-wide SNPs in conglomerate. GCTA uses SNP information to calculate distal genetic resemblance between 'unrelated' individuals. In contrast to family or twin studies that take advantage of assumed patterns of genetic covariance between individuals of differing genetic relatedness, GCTA constructs a genetic relatedness matrix (GRM) between all individuals in the sample that gives an estimate of the genetic similarity between any two individuals due to additive genetic effects (at least those that are tagged by SNPs). Once the GRM has been calculated, the second part of this process involves using a mixed linear model to predict phenotype from the genetic effect of all SNPs. This method has been used to estimate the 'SNP heritability' for numerous traits such as height, body mass index, and intelligence (Davies et al., 2011; Yang, Lee, Goddard, & Visscher, 2011). With regard to smoking phenotypes, GCTA has also been used to estimate the SNP

heritability for smoking initiation (.19), current smoking (.24), and nicotine use and dependence (.18) (Lubke et al., 2012 ; Vrieze, McGue, Miller, Hicks, & Iacono, 2013).

Smoking Phenotype

In addition to attempting to understand the contribution of genetic effects to smoking behavior, multivariate methods have been used to understand the degree to which the genetic and environmental factors that influence each smoking phenotype overlap. For example, smoking initiation is an essential prerequisite to regular use and nicotine dependence, but are the genetic factors that influence initiation the same that influence more persistent forms of smoking, and are there unique factors that contribute to each of these phenotypes? Additionally, how is initiation related to more persistent smoking?

A popular way to investigate the overlapping genetic and environmental contributions to different stages of cigarette smoking is demonstrated by Koopmans, Slutske, Heath, Neale, and Boomsma (1999) for smoking initiation and quantity of cigarettes smoked among adolescents. Three models are used to test three different situations in which either 1) there is a single dimension of liability for both initiation and quantity (thus the genetic and environmental factors are completely shared), 2) the genetic and environmental factors that contribute to liability for initiation and quantity are completely independent of each other, or 3) a 'combined' model that postulates two different liability dimensions for initiation and quantity, but allows for individuals to be non-smokers through either the initiation dimension or by being extremely low on the quantity dimension. Koopmans et al. found support for the combined model; therefore, the genetic and

environmental components that may increase one's liability to smoking could be directly related to initiation itself, or actually be more related to the quantity of cigarettes smoked dimension. Vink et al. (2005) found similar results for smoking initiation and nicotine dependence in an adult sample. Liability to smoke could be determined by genetic and environmental factors related to initiation or those related to the liability dimension of nicotine dependence. Investigating the relationship between the smoking phenotypes in a different way, Maes et al. (2004) examined 3 phenotypes, tobacco initiation, regular tobacco use, and nicotine dependence. Consistent with the models put forth by Koopmans et al. and Vink et al., there was significant overlap in the genetic factors contributing to liability to these three phenotypes. There were also unique genetic factors related to regular tobacco use and nicotine dependence. Further making a case for the role of unique genetic factors related to smoking phenotypes beyond those related to initiation, Madden et al. (1999) found low shared genetic variance between smoking initiation and persistence. Hardie, Moss, and Lynch (2006), too, found low genetic correlations between age of smoking onset and three additional smoking phenotypes: persistence of smoking, cigarettes smoked per day, and duration of smoking cessation.

Gender and Smoking

Additionally, there are known gender differences in the prevalence of cigarette smoking, with a lower percentage of women reporting themselves as current smokers than men (Centers for Disease Control and Prevention (CDC) 2012). There are also reports that the relative contribution of genetic factors related

to smoking may differ depending on gender. However, the findings are mixed; for example, Hamilton et al. (2006) found drastic differences in the heritability of smoking initiation for females (~32%) and males (~71%); yet these differences did not occur for persistent smoking. Contrastingly, in an Australian sample, heritability estimates for initiation were higher in females (67%) compared to males (33%) (Heath et al., 1993).

Cohort Differences in Smoking

Further, over the last seventy years, as the negative health consequences of smoking have become increasingly clear, the social acceptability and prevalence of smoking have changed dramatically (Lopez, Collishaw, & Piha, 1994; National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health 2014). At the beginning of the 20th century, cigarette smoking prevalence was low, but increased across a period when little information was available on the health detriments of smoking. However, with a rise in smoking related mortality and awareness of the harmful effects of cigarette smoking (e.g. the 1964 Surgeon General's Report), prevalence decreased and smoking switched from being seen as socially acceptable to an activity with known negative health consequences (Lopez et al., 1994; Pacheco 2011).

GE Interactions and Cohort

In addition to main effects of the environment and genes on smoking, the two factors may interact, such that the effects of one depend on the other. Studies investigating such gene-environment interactions (GE interaction) on smoking using twin designs have typically investigated what we term *quantitative* gene-

environment interactions, which occur when the magnitude of heritability differs as a function of the environment. For example, attenuated heritability was found for smoking and drinking in individuals who reported high levels of religious adherence (Koopmans, Slutske, Baal, & Boomsma, 1999; Timberlake et al., 2006), and heritability for daily smoking was highest in schools where popular students were smokers (Boardman, Saint Onge, Haberstick, Timberlake, & Hewitt, 2008). With respect to GE interactions across birth cohorts, there was little evidence for a change in heritability for males born in the early 20th century through the 1960s, but for women, heritability was found to increase as prevalence increased (Heath et al., 1993; Kendler, Thornton, & Pedersen, 2000). Boardman et al. found genetic influences for regular smoking for those born in the 1920s, 1930s, and 1950s, but negligible influences for those born in the 1940s and 1960s (Boardman, Blalock, & Pampel, 2010). Finally, Vink and Boomsma (2011) found no difference in the heritability of ever smoking for 18 to 25 year olds in 1993-1995 versus 2009-2010. These studies demonstrate the potential for heritability estimates to differ by birth cohort, but because twins are the same ages, they are blind to whether the actual genetic factors responsible for smoking have changed across environments and/or cohorts.

Qualitative GE interactions occur when the genetic factors that influence traits, or effect of those factors (Carey 1988), change across different levels of the environment. It is important to recognize that a genetic correlation between the same trait across different levels of the environment also measures the extent to which genetic effects are the same or different across environments. Therefore, a

genetic correlation is merely a reparameterization of a qualitative gene-by-environment interaction, and both approaches can be used to estimate the same conceptual construct.

With respect to smoking initiation, it is possible that genetic factors associated with conformity may also be associated with smoking in a society where smoking is commonplace; in social environments where smoking is taboo, a partially different set of genetic factors, perhaps related to deviant behavior, may be related to smoking. Given that the social acceptability of smoking has changed over time, it is possible that the genes or genetic effects associated with smoking initiation have changed across time--i.e., smoking initiation may show a qualitative gene-by-age interaction (or equivalently, a low genetic correlation across different ages). There are some suggestions in the literature that this may be the case. Two extended twin family designs that estimated effects of environmental parental influences on children found modest evidence for negative cultural transmission: whereas genetic factors increased similarity between parents and offspring, parental smoking itself may make their offspring less likely to initiate (Boomsma, Koopmans, Van Doornen, & Orlebeke, 1994; Maes et al., 2006) . However, due to age differences between parents and offspring, qualitative gene-by-age interactions would also reduce parent-offspring resemblance and could appear as negative vertical transmission in extended twin family models.

Although such qualitative gene-by-age interactions/genetic correlations are impossible to detect in twin-only designs (due to lack of variation in age differences between twins) and difficult to detect in family designs (due to confounding with

negative vertical transmission, as explained above), they can be investigated using measured genetic data (Hartz et al., 2012; Vrieze et al., 2012). Given that smoking initiation is likely to be polygenic, with no single variant or set of variants explaining a sizable proportion of the variation (Tobacco and Genetics Consortium 2010), there is a need to determine whether aggregate genome-wide effects on smoking behaviors tend to be the same or different across age.

Summary

In summary, the literature described above makes a case for the following: (1) peers play a role in the cigarette smoking behavior of adolescents, but it is unclear through what mechanism, (2) as evidenced from family studies, genetic factors are also important in smoking behavior, but analyses that have attempted to 'recapture' the genetic contribution from genome-wide data have been unable to account for the genetic variance estimated in twin studies, (3) the contribution of genetic factors may depend on the smoking phenotype in question, (4) there appears to be both overlapping and unique genetic factors that contribute to the different stages of smoking behavior, and (5) the etiology of cigarette smoking may depend on birth cohort. Thus, the above motivates a number of additional queries that we hope will elucidate the etiology of cigarette smoking.

<div align="center">Aims</div>

Here we further investigated the previously described contributors to cigarettes smoking through four studies.

In Study 1, "Adolescent Peer Choice and Cigarette Smoking: Evidence of Active Gene-Environment Correlation?" we explored one possible reason why

individuals resemble their friends' smoking behavior: individuals may be selecting friends who resemble their own smoking habits or homophily. In this study, we were specifically interested in individuals choosing friends who are genetically similar with respect to smoking behavior. If this genetically-based homophily is accompanied by an environment of peer influence, active gene environment correlation will be induced. Thus, we examined the potential for gene-environment correlation in smoking through a twin method implemented by Loehlin (2010) for adolescent alcohol use. Study 2, "Peers And Cigarette Smoking: Catalyst To Initiation Or Precursor Of Regular Use?" extended upon Study 1 by (1) using a more recent cohort of participants (Add Health twins) and (2) examining the relationship between peers and smoking as function of two different smoking phenotypes: experimentation and regular smoking.

Deviating from the twin approaches utilized in Studies 1 and 2, Studies 3 and 4 used whole genome SNP data from unrelated individuals. In Study 3, we examined two smoking phenotypes, age of smoking onset and cigarettes smoked per day. First, we estimated the degree to which the conglomerate of genome-wide SNPs contributed to variance in each of these phenotypes (SNP heritability). We then conducted bivariate GCTA analyses (Lee, Yang, Goddard, Visscher, & Wray, 2012) to determine the degree to which genetic factors responsible for smoking onset overlap with those for smoking quantity or the genetic correlation between onset and quantity. Further, for each phenotype, we examined the SNP heritability for males and females and the genetic correlation between genders. In Study 4, we focused on smoking initiation and used similar methods to estimate the SNP

heritability for two birth cohorts. Additionally, we estimated the genetic correlation between these birth cohorts, determining to what degree individuals born in different eras share the genetic factors that play a role in smoking initiation.

CHAPTER II

STUDY 1: ADOLESCENT PEER CHOICE AND CIGARETTE SMOKING: EVIDENCE OF

ACTIVE GENE-ENVIRONMENT CORRELATION?

Background

We know that peers tend to resemble each other in terms of smoking

behavior, but it is not known why. In addition to pure influence or selection, we

propose a way in which these two forces may act in conjunction. Individuals may

assort or choose friends that are similar to themselves (selection). This chosen peer

group then creates an environment conducive toward smoking or non-smoking

(influence). If individuals make the original assortment based on genetic factors

related to smoking, this produces a correlation between genes and environment. In

Study 1, we examined evidence for gene-environment correlation in cigarette

smoking by replicating the Loehlin (2010) analysis on alcohol use.

Methods

Participants

The sample consisted of 850 twin pairs (514 MZ and 336 DZ same-sex twins)

that participated in the 1962 National Merit Scholarship Qualifying Test as high

school juniors (Loehlin & Nichols, 1976). Exclusions and missing values reduced this

sample to 509 MZ twin pairs (216 male and 293 female) and 330 same-sex DZ twin

pairs (135 male and 195 female).

Items of Interest and Scoring

Zygosity assignment was made on the basis of a questionnaire on reported similarities of a twin pair (Loehlin & Nichols 1976). The original survey contained 3 questions on smoking:

(1) How much do you smoke?

With responses: (1) Never smoked, (2) Used to or occasionally smoke, (3) 1 to 19 cigarettes a day, and (4) greater than 20 cigarettes a day

(2) If you smoke do you inhale the smoke into your lungs?

With responses: (1) Don't smoke, (2) Rarely or never inhale, (3) sometimes inhale, and (4) usually inhale

(3) (How often have you) smoked a cigarette or cigar before breakfast?

With responses: (1) Frequently, (2) Occasionally, and (3) Not at all

Participants reported on the frequency of the said action, and a composite score was assigned for each individual in the following manner. Individuals who had never smoked were given a score of 1 (63% of individuals); individuals who occasionally smoked, but had never inhaled were given a score of 2 (16% of individuals); current or former smokers who had inhaled were given a score of 3 (18% of sample); current or former smokers who had both inhaled and had smoked before breakfast were given a score of 4 (3% of sample).

For the initiation part of our analysis, the composite score was used to dichotomize individuals into categories of initiation status: a binary variable of having never smoked (composite score of 1) or having initiated smoking behavior (composite score > 1). To assess smoking persistence, we only included twin pairs

where each twin had initiated, and persistent smoking behavior was analyzed using the composite score (scores 2 though 4) described above.

Individuals missing scores on any of the above items were assigned scores based on responses to the other smoking items. For example, a respondent who omitted an answer to "how much do you smoke" but reported smoking before breakfast was assigned a 4. Such assignments involved only a small proportion of the sample (1.50%) and were made without knowledge of zygosity and twin's smoking status.

For the measure of shared friends, participants were asked: "Do you and your twin have the same or different friends?" Responses were on an ordinal scale ranging from a score of 1 (all shared friends) to 4 (few to no shared friends). As in Loehlin (2010), we reverse scored this item for ease of interpretation. Thus, a score of 1 indicated few shared friends between twins and a score of 4 indicated complete sharing of friends between twins. To get a shared friends score for each twin pair, we averaged the two twins' shared friend scores. In the case of a missing shared friend score for one twin, we assigned that twin the score given by the other twin in the pair. If both twins had no score, this pair was excluded from the shared friends portion of the analysis.

Evaluation of Heritability and Gender Differences

All four gender by zygosity groups were fitted to a multifactorial threshold model, which assumed a threshold imposed on an underlying continuous distribution of factors related to liability of the smoking composite. The first question of interest was whether there was a sex difference in the additive genetic

and environmental influences of the composite smoking variable that may account for differences in results for male and female twin pairs. Here, we tested whether the parameters of an ACE model could be equated across genders.

To examine the relationship between smoking and number of friends in common, we followed Loehlin (2010) and correlated the absolute value of the difference in pair smoking scores with the measure of common friends. This was done for the composite, initiation, and persistent smoking measures.

All statistical and model fitting analyses were conducted in R version 2.12.0 and OpenMx version 1.0.3-1505 (Boker et al., 2011).

<div align="center">Results</div>

For simplicity, we report descriptive and univariate heritability measures for the composite smoking score that includes both components of the initiation and persistence measures.

Gender Differences in Shared Friends and Composite Smoking

Tables 2.1 and 2.2 present descriptive statistics for the gender-by-zygosity groups. Gender and zygosity effects were tested using a two-by-two ANOVA with an interaction term. No interactions were significant, so here we report the marginal differences.

Males (M = 1.68, SD = .80) had higher smoking composite scores than females (M = 1.57, SD = .66), but the difference was only marginally significant ($t(837)$ = 1.95, p = .05). Females (M = 3.14, SD = .65) reported more shared friends than males (M = 3.02, SD = .59), ($t(836)$ = -2.67, p = .008).

Table 2.1

Means and standard deviations on smoking composite and absolute difference in smoking composite between twins for the four samples.

		Composite Smoke Score		Absolute Difference in Smoke Score	
	N	Mean	*SD*	Mean	*SD*
MZ Males	216	1.62	.78	.40	.65
MZ Females	293	1.53	.80	.25	.55
DZ Males	135	1.77	.82	.53	.73
DZ Females	195	1.62	.79	.47	.78

Table 2.2

Mean and standard deviations for mean shared friends for the four samples.

	N	Mean	*SD*
MZ Males	215[a]	3.17	.53
MZ Females	293	3.27	.60
DZ Males	135	2.79	.60
DZ Females	195	2.94	.66

Zygosity Differences in Shared Friends and Smoking

Across sex, MZ and DZ twin pairs only marginally differed in their smoking composite scores, with DZ twin pairs (*M* = 1.68, *SD* = .81) having slightly higher composite scores than MZ twin pairs (*M* = 1.57, *SD* = .79), (*t*(837) = -2.01, *p* = .05). However, DZ twin pairs (*M* = .49, *SD* = .76) were significantly more divergent in their smoking behaviors (i.e., absolute difference scores) than MZ twin pairs (*M* = .31, *SD* = .60), (*t*(837) = -3.62, *p* = .0003). MZ twin pairs (*M* = 3.22, *SD* = .57) also shared significantly more friends than DZ twin pairs (*M* = 2.87, *SD* = .64), (*t*(836) = 8.03, *p* < .001).

Heritability of Smoking Composite

Table 2.3 presents the biometrical genetic model for the smoking composite measure. Thresholds could be equated across twin pair ($\Delta\chi^2$ (8) = 3.19, *p* = .92) and zygosity ($\Delta\chi^2$ (4) = 6.23, *p* = .18), but not gender ($\Delta\chi^2$ (2) =12.95, *p* = .002). The ACE model with equal parameters across gender did not significantly reduce fit compared to the model that allowed these parameters to vary separately for each sex ($\Delta\chi^2$ (3) = 4.61, *p* = .20). Therefore, analysis was continued jointly for male and female twin pairs. Both the Akaike Information Criterion (AIC) and the likelihood ratio tests suggested that the model containing all three variance components is to be preferred. Compared to a full model, the CE model ($\Delta\chi^2$(1) = 14.46, *p* < 0.001) and AE model ($\Delta\chi^2$(1) = 8.98, *p* = .003) could convincingly be rejected.

Table 2.3

Smoking composite model fit statistics and variance component estimates with 95% confidence intervals for ACE and nested AE and CE models.

Model	EP	df	-2LL	AIC	Δχ²	Δdf	p	VA	VC	VE
ACEª	7	1623	2594.27	-651.73	---	---	---	.43 (.20-.69)	.39 (.15-.60)	.18 (.13-.24)
AE	6	1624	2603.25	-644.75	8.98	1	<.01	.83 (.78-.87)	---	.17 (.13-.22)
CE	6	1624	2608.73	-639.27	14.45	1	<.01	---	.74 (.68-.79)	.26 (.21-.32)

a. Threshold values: Males thresholds: t1: .25, t2: .54; Female thresholds: t1: .48, t2: .43

Differences Between MZ and DZ Correlations

To evaluate the possibility of active rGE, we examined the polychoric correlation between absolute differences in smoking between twins and the mean amount of shared friends each twin reported (Table 2.4).

Table 2.4

Polychoric correlations between absolute difference in twin pair smoking behavior and average shared friends for composite, initiation, and persistence smoking measures.

	Composite	Initiation	Persistence
Group	r ª (s.e)	r ª (s.e.)	r ª (s.e.)
MZ males	.01 (.09)	-.05 (.10)	-.07 (.16)
MZ females	-.10 (.09)	-.02 (.10)	-.25 (.15)
DZ males	-.23* (.10)	-.39* (.11)	.16 (.20)
DZ females	-.29* (.09)	-.31* (.09)	-.10 (.19)

* p < .05
a. All correlations met assumption of bivariate normality.

For the smoking composite, both the correlations for MZ males and MZ females did not significantly differ from 0. However, the DZ correlations were significant for both males and females. Hence, for DZ but not MZ twins, more shared friends predicted greater twin similarity in smoking. However, the difference between the MZ and DZ correlations was not significant for female ($\Delta\chi^2$ (1) = 2.44, p = .12), but marginally significant for male ($\Delta\chi^2$(1) = 3.49, p = .06) twin pairs.

For smoking initiation, polychoric correlations between shared friends and differences in smoking for MZ males and females did not significantly differ from 0. Yet, for DZ males and females, the correlation between shared friends and absolute difference in smoking initiation was significant. Further, for both males ($\Delta\chi^2$(1) = 6.15, p = .01) and females ($\Delta\chi^2$(1) = 4.48, p = .03) the DZ correlation between smoking and shared friends was significantly stronger than the MZ correlation. Therefore, the initiation phenotype gave the same results as the composite measure.

For smoking persistence, however, the previous pattern of correlations was not observed. Correlations for MZ males (n =61), MZ females (n = 76), DZ males (n =44), and DZ females (n =48) did not significantly differ from 0. Thus, shared friends did not predict similarity in smoking status beyond initiation.

Discussion

From the results, two major findings are highlighted. First, in DZ, but not MZ, twin pairs there was a relationship between number of shared friends and similarity of the smoking composite and initiation score. Thus, these results were consistent with the possibility that genetic differences within DZ twin pairs may influence the choice of friends with characteristics that correlate with each twin's unique genetic

predispositions, which in turn may explain corresponding differences in the pair's smoking.

Second, when the smoking variable was reduced to include only individuals that had initiated for the smoking persistence part of the analysis, we did not find this same pattern. This was unsurprising especially given the small sample size of roughly between 40 to 80 twin pairs per group. The two items used to define smoking persistence, whether one inhales or smokes before breakfast, may also not have been the most optimal or relevant measure of smoking persistence in an adolescent sample (Heatherton, Kozlowski, Frecker, & Fagerström, 1991). Therefore, our composite smoking measure may more accurately assess initiation rather than smoking persistence, and while our results for initiation may be valid, limitations of the current dataset may be unable to conclude on the effect of shared peers on smoking persistence past the stage of initiation.

In conclusion, we have evidence for homophily, or that individuals are choosing friends based on their own genetic predispositions with regard to smoking behavior. They thus may be selecting into environments that, through the possibility of further peer influence, will further promote expression of their predisposed smoking preferences. However, before substantive interpretation of these findings, it is important to first rule out other mechanisms that could contribute to our observed pattern of correlations.

One potential issue is the possibility that peer influence may violate a cardinal assumption of the twin method, namely, that the correlations in latent, trait-relevant environmental values are equal for MZ and DZ pairs. Similar peer

groups for MZ twins may partly arise from the extra attention these twins receive by being together in a group. Kendler and Gardner (1998) evaluated whether this mechanism could play a role for both smoking initiation and nicotine dependence. They reported that twins with higher 'co-socialization' scores, a factor based upon items related to how often twins socialized together, resembled each other more with respect to smoking initiation but not nicotine dependence. However, this mechanism would also predict high correlations between peer group differences and within-pair smoking differences for *both* MZ and DZ twins. Thus, our results are inconsistent with this being a *strong* mechanism behind differences in correlations between MZ and DZ twins.

A second mechanism that could influence the relationship between peers' smoking behavior is passive assortment, or peer associations based on background variables that are correlated with smoking. Family socioeconomic status is a clear example. In nationally representative twin samples, simple geography, ethnicity, culture and religious affiliation may influence both peer similarity and smoking behavior (Degenhardt, Chiu, Sampson, Kessler, & Anthony, 2011). However, neither Loehlin's nor our results were consistent with pure passive assortment as the pre-eminent mechanism for the correlation between self- and peer-drug use behaviors. Friendship groups based on background factors correlated with smoking should be the same for MZ and DZ twins, leading to identical correlations for MZ and DZ twins in peer group differences and within-pair smoking differences.

Hence, the most likely factors contributing to the observed homogeneity in peer group smoking are a combination of homophily and, possibly, peer-influence.

Unfortunately, we could not quantify the precise contribution of each mechanism with the current data, because there were no data on smoking in the twin's friends. The data, however, are not consistent with a strong role for peer influence. If peer influence were very important, then we should have observed at least a trend toward significance in the correlations between smoking status and peer-group similarity in MZ twins. Yet, with the exception of smoking persistence, all MZ correlations in Table 2.4 are very close to 0.

Hence, the pattern of results definitely supported homophily as an important mechanism. In contrast to peer influence alone, homophily predicts that peer groups should be more similar for MZ than DZ twins. Hence, our results *could* be consistent with active rGE as an explanation why individuals and their peer groups tend to share smoking habits, if this homophily is accompanied by peer influence. However, we cannot definitively conclude that both of these requirements for rGE are taking place given the limitations of a cross-sectional, twin dataset. Future analysis on more expanded datasets that include twins, adoptees, and siblings raised apart may illuminate the specific contribution of rGE, especially with longitudinal data from adolescence until early adulthood.

A further caveat was the age of the participants. Participants were only evaluated at the single time point of juniors in high school (~17 years of age). However, longitudinal analyses have suggested that genetic and environmental components related to peer influences might vary across different ages within the span of early adolescence to young adulthood (Vink et al., 2003; White et al., 2003).

Additionally, given only a single time point of evaluation, it is important to take into consideration cohort differences in the etiology of smoking behavior. Boardman et al. (2010) demonstrated the dynamic nature of heritability for regular smoking behavior across a series of cohorts born in the United States, finding rather negligible genetic influences for those born in the 1940s, the cohort of the National Merit Twins (Loehlin & Nichols, 1976). They reasoned that social pressures might have pushed the popularity of smoking to a level in which genetically vulnerable individuals were no more likely to smoke than those without genetic predispositions toward smoking. Kendler et al. (2000) examined the heritability of regular smoking by birth cohort for males and females separately in Sweden and found that heritability for women was actually greatest for those born after 1940. The reasoning in this case was not so different from Boardman et al.'s, but appears to have quite a different effect on the manifestation of genetic influences for women: decline in the social stigma of women smoking may have allowed women to partake in behaviors aligned with their genetic propensities. Given evidence of these changes in the etiology of smoking behavior over time, it becomes difficult to generalize both the genetic and peer influence components of our model to the present era. Therefore, given that we only had participants at a single age from a single point in time, it is possible that our results illustrated only a snapshot of the true mechanisms by which smoking behavior may be regulated through the influence of genes on peer choice.

Further issues discussed by Loehlin (2010) regarding this sample included the ability of the questionnaire to accurately assess the number of shared friends

and the key behavior (smoking in this study). Also addressed was the fact that variance in the key behavior may have been restricted by a select sample of high-achieving participants.

Thus, our findings are consistent with Loehlin's (2010) for alcohol-related behavior in that there appeared to be a pathway to smoking through both genes and peer groups. However, the contribution of active rGE that requires both genetically-based homophily and additional peer influence remains unclear. Yet, given evidence in the literature for the dual contribution of selection and peer influence to smoking behavior (Simons-Morton & Farhat, 2010), rGE might be a likely scenario. Thus, future investigations should specifically test for the contribution of rGE using datasets more amenable to quantifying the alternative contributions of unequal twin environments, passive assortment, or peer influence/homophily that may produce similar results.

In conclusion, our results further evidenced the contribution of genetics on exposure to environments that may influence our behaviors. However, given the unclear effect of sex and age on such findings as well as the inconclusive evidence as to the full range mechanisms at work, the next step is to further illuminate both the ways by which rGE might specifically be at work and what other factors may contribute to homogeneity between individuals and their peer groups in smoking behavior. Particularly, gaining a clearer understanding of this association will allow future research to examine ways toward effective prevention of the problems associated with cigarette smoking.

CHAPTER III

PEERS AND CIGARETTE SMOKING: CATALYST TO INITIATION OR

PRECURSOR OF REGULAR USE?

Background

We know that peers play an important role in cigarette smoking, but we do

not know at what stage of the smoking process, from initiation to regular use, that

peers are the most salient factor. In Study 2, we expanded upon the first study, using

a more recent dataset, to again evaluate the potential for gene-environment

correlation for smoking. However, we also focused on whether this mechanism

plays a differential role based on whether cigarette smoking is defined as smoking

experimentation or smoking cigarettes regularly.

Method

Participants

Our participants were from Waves 1 (W1) and 2 (W2) of the National

Longitudinal Study of Adolescent Health, a representative study of American

adolescents in grades 7-12, with the first wave of data collected in 1994-95 (Harris

2011). Specifically, we focused on 754 twin pairs collected as a part of the "genetic

pairs data" that includes 3000 pairs with varying degrees of genetic relatedness. The

twin sample included five gender and zygosity groups: identical (MZ) male and

female twin pairs, fraternal (DZ) male and female twin pairs, as well as opposite-sex

fraternal twin pairs.

Measures

In this analysis we focused on two stages of smoking: experimentation and regular use.

Smoking Experimentation. To measure smoking experimentation we constructed a score from items related to trying cigarettes. For each individual, we assigned a score based on his or her endorsement of each of the items as follows for W1:

Score:

0	:	No smoking items endorsed.
1	:	Had a puff of a cigarette, but not a whole cigarette.
2	:	Smoked a whole cigarette.

Because W2 did not include an item about smoking a whole cigarette, the W2 smoking initiation variable was dichotomous (0 or 1), in which the participant either did or did not have a puff of a cigarette.

Regular Smoking. To focus on factors separating those that smoke regularly from those having only experimented with smoking, we restricted our sample to twin pairs where both members had reported at least having tried cigarettes. For each individual we then computed a score from the 1st principle component of the following items for W1:

1. Have you ever tried cigarette smoking, even just 1 or 2 puffs?
2. How old were you when you smoked a whole cigarette for the first time?
3. Have you ever smoked cigarettes regularly, that is, at least 1 cigarette every day for 30 days?
4. How old were you when you first started smoking cigarettes regularly (at least 1 cigarette every day for 30 days)?
5. During the past 30 days, on how many days did you smoke cigarettes?

6.   During the past 30 days, on the days you smoked, how many cigarettes did you smoke each day?

Similar questions were used to compute each participant's component score at W2.

Shared Friends. As a measure of mutual friends, each twin was asked about their co-twin, "How much time do you spend with the same friend or group of friends?" Twins could respond with: "1: a lot", "2: some", "3: little", or "4: none." Because higher numbers originally indicated a lower index of shared friends, we reverse scored this item for ease of interpretation; therefore higher numbers indicated more mutual friends, with a score of 0 indicating no shared friends. When we had responses from both twins, we averaged this value to get a value for each twin pair. In cases where we only had a single twin's response, we used that twin's given value for the pair.

Analyses

For smoking experimentation and regular smoking, at each wave of data collection, the twin smoking correlation was estimated. For smoking experimentation (an ordinal variable at W1 and a dichotomous variable at W2), we fit a multifactorial threshold model to the data, which assumed a threshold imposed on an underlying continuous distribution of factors related to liability of smoking initiation. Thus, we estimated the polychoric (W1) or tetrachoric (W2) correlation (r) between twin 1 and twin 2's smoking experimentation score. For regular smoking, a continuous variable, we estimated the intraclass correlation between twin 1 and twin 2's smoking score, and, due to random assignment to 'twin 1' or 'twin 2' status, double entered twins and corrected for the degrees of freedom in our analysis.

To match the validity of our measured variables with respect to heritability estimates pertaining to smoking experimentation and regular use in the literature, we fit a univariate twin model to each smoking phenotype at each wave. We used the raw data from the previous correlations described above and included additive genetic (A), shared environmental (C), and unique environmental or measurement error (E) components. Subsequent sub-models that dropped each component of the ACE model were fit and compared to the original ACE model by the significance of the difference in chi-square between the full and reduced model. Only complete twin data for the smoking variables was used.

The models described above estimated the twin smoking correlation (α), irrespective of shared friends:

$$r_{twin1,twin2} = \alpha \qquad\qquad (1)$$

However, in order to assess the degree to which shared friends impacted twin smoking similarity we fit a second model that estimated the twin correlation as a function of two parameters, α and β:

$$r_{twin1,twin2} = \alpha + \beta \text{ (Average Shared Friends)} \qquad (2)$$

In this case, the α parameter is the y-intercept or twin smoking correlation when twins spend time with none of the same friends, and the β parameter is the degree to which the shared friends variable influences the twin smoking correlation. In this case, we tested whether the β parameter could be equated across twin group. In other words, we tested whether β was significantly different for MZ versus DZ twins. To justify inclusion of the β parameter, we tested the significance of the difference between the two models that either excluded or included the β

parameter. Identifying the most parsimonious model with respect to unique β terms was done by testing whether reduction of each β term resulted in a significant detriment to the fit of the model as defined by a significant difference in chi-square values between the two competing models.

Again, all models were fitted to raw data and analysis was completed using OpenMx version 1.2 (Boker et al., 2011) and R version 2.12.0.

Results

Twin correlations are reported in Table 3.1 (W1) and Table 3.2 (W2). In all cases, there were no statistically significant gender differences in the twin correlations or differences between correlations of DZ same-sex and DZ opposite-sex twin pairs. Therefore, analyses were continued on the combined MZ and DZ twin groups.

Table 3.1

Intraclass Polychoric Twin Correlations: Experimentation

	MZ		DZ	
	N	r (s.e.)	N	r (s.e.)
W1				
Combined	297	.72 (.04)	441	.55(.05)
Females	146	.74 (.06)	117	.53 (.10)
Males	151	.70 (.06)	127	.62 (.08)
Opposite Sex			197	.48 (.08)
W2				
Combined	269	.62 (.07)	397	.43 (.07)
Females	134	.64 (.09)	112	.59 (.11)
Males	135	.61 (.10)	109	.46 (.13)
Opposite Sex			176	.30 (.11)

Table 3.2

Intraclass Twin Correlations: Regular Use (Only Smokers)

	MZ		DZ	
	N	r (s.e.)	N	r (s.e.)
W1				
Combined	133	.69 (.03)	165	.43 (.05)
Females	66	.70 (.05)	41	.35 (.10)
Males	67	.69 (.05)	60	.37 (.08)
Opposite Sex			64	.52 (.09)
W2				
Combined	73	.81 (.03)	88	.38(.07)
Females	42	.78 (.04)	25	.33 (.13)
Males	31	.84 (.04)	31	.34 (.11)
Opposite Sex			32	.49 (.14)

We report estimates of the variance components from a univariate twin analysis for each of these variables in Table 3.3. Both additive genetics and shared environmental factors contributed to twin resemblance in our smoking experimentation variable at W1. However, though there was a familial effect, it was unclear whether this familial effect in the full model could be reduced to include either additive genetic or shared environmental influences for smoking initiation at W2. For regular use, both additive genetic and shared environmental influences contributed to regular smoking at W1, though the magnitude of additive genetic factors was stronger than that of the shared environment. At W2, only additive genetic and non-shared environmental factors influenced regular smoking.

Table 3.3

Univariate Twin Analysis

Smoking Phenotype	Model	EP	-2LL	AIC	$\Delta\chi^2$	Δdf	p	VA	VC	VE
Experimentation W1[a]	**ACE**	**5**	**2834.06**	**-123.94**	**---**	**---**	**---**	**.41**	**.33**	**.26**
	AE	4	2842.36	-117.64	8.30	1	<.01	.77	---	.23
	CE	4	2843.52	-116.49	9.45	1	<.01	---	.62	.38
Experimentation W2[b]	ACE	4	1749.82	-938.18	---	---	---	.39	.24	.37
	AE	**3**	**1752.09**	**-939.91**	**2.28**	**1**	**.13**	**.66**	**---**	**.34**
	CE	**3**	**1753.58**	**-936.42**	**3.77**	**1**	**.05**	**---**	**.51**	**.49**
Regular Use W1	**ACE**	**5**	**1517.32**	**335.32**	**---**	**---**	**---**	**.40**	**.27**	**.33**
	AE	4	1521.30	337.30	3.97	1	0.05	.68	---	.32
	CE	4	1525.71	341.71	8.39	1	<.01	---	.57	.43
Regular Use W2	ACE	5	865.83	231.83	---	---	---	.85	-.03	.18
	AE	**4**	**865.87**	**229.87**	**0.04**	**1**	**.85**	**.81**	**---**	**.19**
	CE	4	892.91	256.91	27.07	1	<.01	---	.58	.42

Bolded line indicates best fitting model
[a] Thresholds equated across zygosity; $\Delta\chi^2$ = 2.61.; ΔDF = 2; p = .27
[b] Thresholds equated across zygosity; $\Delta\chi^2$ = 1.3 x 10^-6; ΔDF = 1; p = .99

Effect of shared friends

  Change in chi square estimates and their significance for the specified models as well as parameter estimates for the best fitting models are listed in Table 3.4.

  Smoking Experimentation. At W1, starting with a saturated model that allowed for zygosity specific thresholds, the β parameter, or effect of shared friends, could be equated for MZ and DZ twins. However, dropping this parameter from the model produced a marginal, but not statistically significant, detriment in model fit. At W2, the effect of shared friends could also be equated across zygosity, and dropping this parameter produced a significant decrease in model fit. Therefore, for smoking initiation, predominantly at Wave 2, the β parameter or effect of mutual friendships on the twin smoking correlation, could not be excluded.

  Regular Smoking. At W1, the β parameter could not be dropped from the model without significantly reducing model fit. It could, however, be equated across all gender and zygosity groups to get a single β parameter, indicating an association between twin similarity for regular smoking behavior and time spent with the same group of friends. However, the β parameter at W2 could not be equated across zygosity, resulting in different MZ and DZ values. While the MZ β value could be dropped from the model without a significant detriment in model fit, dropping the DZ β value decreased fit of the model. Therefore, increased time spent with the same group of friends increased the DZ, but not MZ, twin correlation.

Table 3.4.

Effect and estimates of the shared friends parameter (β)

	$\beta_{MZ} = \beta_{DZ}$[a]	$\beta = 0$[b]	β_{MZ}	β_{DZ}
	$\Delta\chi^2$ (1)	$\Delta\chi^2$ (1)		
Experimentation W1	0.46	3.71[#]	.08 (.04)	.08 (.04)
Experimentation W2	0.35	6.43*	.17(.07)	.17(.07)
Regular Use W1	1.05	60.26***	.22 (.03)	.22(.03)
Regular Use W2	4.31*	MZ: 2.79	-.10 (.01)	.23(.10)
		DZ: 6.94**		

[#] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Saturated model included $\alpha_{MZ}, \alpha_{DZ}, \beta_{MZ}, \beta_{DZ}$, and thresholds $_{MZ/DZ}$ (Experimentation)
[b] Saturated model included $\alpha_{MZ}, \alpha_{DZ}, \beta$ (based on whether $\beta_{MZ} = \beta_{DZ}$), and thresholds $_{MZ/DZ}$ (Experimentation)

Discussion

When smoking was defined as smoking experimentation, the degree to which twins spent time with the same group of friends affected their resemblance for smoking experimentation, especially at an older age (W2). When the definition of smoking was altered to refer to regular use, that included items related to smoking frequency and quantity, more time spent with the same peer group increased twin resemblance for smoking behavior. However, this effect only persisted for fraternal twins at the second wave of data collection.

The goal of this analysis was to examine the differential impact of peer relationships at different stages of the smoking trajectory. Peer relationships are a known contributor to smoking status (Kobus, 2003), but the details of exactly how these relationships impact adolescent choices to initiate and continue smoking are unclear.

Our univariate analyses echoed findings from the literature implicating a genetic component to smoking status. With the exception of the regular use variable

at the second wave of data collection, it appeared that the shared environment also played a role, though the magnitude of these estimates was smaller than those for the additive genetic contribution (Table 3.3). This is an interesting factor to note when considering to what degree the variable mutual peers was a reflection of shared environmental or additive genetic components that may have resulted in twin smoking resemblance.

In Study 1, we examined a similar question in assessing whether twins may share friends as a result of being more genetically similar to each other with regard to factors influencing smoking. If such was the case, there would be a correlation between smoking status resemblance and number of shared friends, for DZ, but not MZ twins, because MZ twins are already at maximal genetic relatedness, while DZ twins vary in their genetic resemblance. For a composite smoking variable that included items related to both initiation and more regular use, we found potential evidence for genetically-based homophily, or that resemblance between individuals and their peers may be due to choosing friends that are more genetically similar to themselves in terms of smoking behavior. This finding was extended to apply to smoking initiation, but not for smoking progression.

While the present analysis did not allow us to assess the specific role that genetically-based homophily may have had in the homogeneity of smoking behaviors within peer group, it did tell us that sharing peer groups increased similarity of smoking behavior, whether that smoking behavior was the initial first cigarette puff or more regular smoking behavior. At this stage, the relative role of shared environments and shared genes on friendship similarity in twins remained

unclear. As the effect of shared peers was statistically equivalent for both MZ and DZ twins (with the exception of regular smoking at W2) we could not rule out either of these possibilities, and it may be the case that the variable of shared peers reflected both genetic and environmental sources of resemblance.

However, for regular smoking at W2, we did not see this pattern; shared friends did not impact MZ twin resemblance but increased resemblance of DZ twins. This result for regular smoking at W2 replicated the previously described findings in Study 1 for smoking and Loehlin (2010) for alcohol use. It should also be noted that these two studies assessed twins that were more similar in age to the W2 data in the present analysis. Interestingly, for this variable, we could also exclude the shared environmental component of our univariate model (Table 3.3). Thus, shared friends for regular smoking at this wave of data collection may have been an indicator of increased genetic resemblance among DZ twins, which, as a result, increased smoking behavior resemblance. However, given such a high initial MZ correlation and reduced sample sizes for this variable (Table 3.2), we remain cautious with interpreting this result.

The major limitation of our analysis was that we were still unable to decompose what shared peers among twin pairs tells us in regard to genetic and environmental sources of resemblance in smoking behavior. Future analyses should target this question specifically, as it may provide insight into exactly how peer relationships affect adolescent decisions to smoke. A further limitation is that our shared peer variable was not indicative of whether the shared peers smoked themselves or not. This dataset does, however, have information regarding the

number of friends that each individual reports as being a smoker. Hoffman, Monge, Chou, and Valente (2007), using this same dataset (Waves 1 and 2 of AddHealth), evaluated a longitudinal model that assessed an individual's smoking behavior at two time points (W1 and W2) as a function of the number of friends that smoke at W1 and W2. Their result, finding an association between adolescent smoking at the initial time point and smoking friends at the second time point, reiterates the impact of selection on peer homogeneity in smoking behavior. Yet, they did not find support for peer influence, as the relationship between friend smoking at W1 and individual smoking at W2 was negative in direction. We are cautious about interpreting their results, however, as W1 and W2 are only a year apart, which may not be enough time to develop or assess the dynamics of peer formation and development of smoking habits. Information about the smoking behavior of friends at later time points (not available in this dataset) would be instrumental in clarifying the impact of both selection and peer influence.

We also point out that the individuals in this sample were middle to high school aged at the time of assessment, and, if having even experimented with smoking, might have not yet taken on habits relevant to patterns of regular use by this time. Using data from later waves may have rectified this issue; however, no data on shared friends was available at subsequent waves following Wave 2.

In conclusion, we confirmed previous reports of the role that peer relationships have in cigarette smoking behavior, and uniquely, we demonstrated that this role was relevant to both the experimentation and continued use of cigarettes. We also presented the possibility that similarity within peer groups of

smoking behavior might be due to shared genetic propensities, but we could not rule out environmental factors contributing to this resemblance as well. Understanding the complexity involved in adolescent choices to both initiate and continue to use cigarettes, and particularly how peer relationships impact each of these choices, will be important in the development of effective strategies to curtail the possible adverse effects of these decisions.

CHAPTER IV

SNPS AND SMOKING: WHAT CAN THE AGGREGATE OF GENOME-WIDE SNPS TELL

US ABOUT GENETIC LIABILITY TO SMOKING INITIATION AND QUANTITY

SMOKED?

Background

We know that genetics is a relevant factor in cigarette smoking behavior, but studies that focused on single or small groups of variants within whole-genome data have yet to recapture the entirety (or any sizable portion) of the genetic variance as estimated from the twin literature (Tobacco and Genetics Consortium, 2010). Here, we used the conglomerate of common SNPs to estimate the 'SNP heritability' for smoking age of onset including initiation and quantity smoked using GCTA (Yang, Lee, Goddard, & Visscher, 2011). GCTA uses genome-wide single nucleotide polymorphism (SNP) data to calculate the genetic similarities between classically 'unrelated' individuals, and determines the degree to which genetic similarities are associated with phenotypic similarities to estimate the additive genetic effects tagged by common SNPs. When standardized by the phenotypic variance, this method estimates the "SNP heritability" (SNP h^2): the proportion of phenotypic variance explained by the aggregate effects of all SNPs.

We also know that different smoking phenotypes may have differing genetic etiologies, but the twin literature has been mixed in regard to the degree that the genetic factors influencing each phenotype overlap. Thus, we also conducted a

bivariate analysis in GCTA (Lee, Yang, Goddard, Visscher, & Wray, 2012) to gain a better understanding of the degree to which the genetic factors for smoking initiation overlap with those that influence quantity smoked.

Method

Participants

Genetic and phenotypic data obtained from dbGaP consisted of 10,162 participants from the Atherosclerosis Risk in Communities (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA) that were genotyped on the Affymetrix 6.0 SNP chip.

The ARIC data was collected from four US communities to prospectively study atherosclerosis and related cardiovascular conditions. As a part of the cohort component of their study, 15,792 individuals aged 45 to 64 were included. Cohorts were selected by probability sampling within each community, and households with age-eligible individuals were selected for home interviews and invited to participate in a clinic examination starting in 1987-89. In the present analysis we used only the data from the first examination. Additional study details have been described elsewhere (Investigators 1989). The MESA data was collected from six US centers starting in 2001. 6,500 individuals between the ages of 45 and 84 were recruited to investigate sub-clinical cardiovascular disease. Details have been reported in Bild et al. (2002).

Quality control procedures were completed initially within each dataset and then again in the merged dataset. Only self reported European Americans were used. We removed ethnic outliers by visually inspecting the first two dimensions of

a multidimensional scaling analysis plot. We also removed individuals with missing phenotypes or covariates, genotyping rates less than 95%, discrepancies between reported and genotypic sex, and with heterozygosity estimates greater than 3 standard deviations from the mean. We removed one of two individuals based on high pairwise genetic relatedness (indexed by $\hat{\pi}$ values >. 05 in GCTA), because closely related individuals, in addition to sharing more genetic variants, may also share more similar environments. The final sample included 8494 individuals.

Measures

    Age of smoking onset. In concordance with Heath, Martin, Lynskey, Todorov, and Madden (2002), we defined smoking initiation in terms of age of smoking onset. We assessed smoking age of onset in both datasets and included all participants for this measure. In ARIC, this variable was assessed by interview with the question, 'How old were you when you first started regular cigarette smoking?' In MESA, age of onset was assessed by questionnaire with the question 'How old were you when you first started smoking cigarettes?' We coded this variable based on increasing liability such that individuals who reported not ever smoking were given a score of 0, individuals that started smoking after age 18 were given a score of 1, and individuals that smoked before age 18 were given a score of 2.

    Quantity smoked. Only individuals who had initiated smoking (scores of 1 and 2 on age of onset measure) were included in the quantity smoked analysis. We assessed quantity through reported cigarettes smoked per day. In ARIC participants were asked "On the average of the entire time you smoked, how many cigarettes did you usually smoke per day?" and MESA participants answered the question, "On

average, how many cigarettes a day do/did you smoke?". Participants reported the

number of cigarettes smoked per day, and this was included as a continuous

variable in the analysis.

Analyses

We used GCTA to estimate the univariate SNP h^2 for both smoking age of

onset using all participants (n = 8494) and quantity smoked for only participants

who had initiated smoking (n = 5075). Further, we examined the overlap of genetic

factors responsible for smoking onset in each gender, and repeated this analysis for

quantity smoked. We then conducted a bivariate analysis that estimated the SNP

correlation (SNP-r_g) between age of onset and quantity smoked, providing an

estimate of the degree to which the SNPs that influence onset are the same as those

that influence quantity smoked. Age, gender, dataset, and the 1st ten ancestry

principal components (to control for the possibility of ethnic stratification (Price et

al., 2006) were included as covariates in all analyses. All analyses and quality

control were conducted using PLINK v1.07(Purcell et al., 2007), PLINK v1.90b2b

(Purcell & Chang, 2014), R version 3.0.1 (Team 2013), and GCTA version 1.24.3

(Yang et al., 2011b).

<div align="center">Results</div>

Onset versus quantity smoked

Common SNPs contributed to both age of smoking onset (SNP h^2 = .08; SE =

.04; p = .02) and quantity smoked (SNP h^2 = .12; SE = .06; p = .03). Additionally, the

SNP correlation between onset and quantity (SNP-r_g = .39; SE = .34; $p_{\text{SNP-rg} = 1}$ = .08; p

$_{\text{SNP-rg} = 0}$ = .14) revealed that there is not complete overlap between the genetic factors related to smoking onset versus smoking quantity.

Gender effects for smoking onset and quantity

When each phenotype was subdivided by gender, SNPs did not significantly contribute to male (n = 4168) smoking onset (SNP h^2 = .12; SE = .08; p = .06), male (n = 2897) smoking quantity (SNP h^2 = .10; SE = .11; p = .16), female (n = 4362) smoking onset (SNP h^2 = .06; SE = .07; p = .18), or female (n = 2178) smoking quantity (SNP h^2 = .07; SE = .14; p = .32). The SNP correlations between males and females for smoking onset (SNP-r_g = .74; SE = .73) and smoking quantity (SNP-r_g = 1; SE = 1.51) were high, indicating substantial overlap between the genetic factors related to smoking onset and quantity smoked for males and females.

Discussion

Here we conclude that the conglomerate of genome-wide SNPs contributes to both smoking onset and quantity smoked. From the bivariate analysis, it also appears that there are both unique and shared genetic factors responsible for smoking onset and quantity. In other words, the set of genetic factors that contribute to age of smoking onset do not entirely account for the amount that an individual smokes post initiation; there appear to be a unique set of factors related to quantity smoked. Evidence for a separate genetic liability related to more persistent smoking beyond the liability to smoking initiation reiterate what has been found in family studies (Hardie et al., 2006; Koopmans et al., 1999; Madden et al., 1999; Maes et al., 2004; Vink et al., 2005).

However, when we divided this analysis by gender, we were unable to account for any variance in either phenotype using the conglomerate of SNPs, and there did not appear to be unique genetic factors related to each gender in terms of smoking onset or quantity. One possibility for our inability to assess these effects by gender may be related to the reduction in sample size and, resultantly, power as a result of dividing the dataset by gender, especially in the bivariate analysis (Visscher et al., 2014). Previous analyses have drawn attention to the importance of including gender as a factor when examining the genetics of smoking phenotypes. For example, while not finding large differences in the actual heritability for each gender, Morley et al. (2007) did find that genetic correlations among three smoking phenotypes (onset, consumption, and persistence) was higher in males than females, and that environmental connections between these traits may be higher in women. Thus, while gender may play a substantial role, the present analysis was unable to draw any meaningful conclusions regarding how genetic factors related to onset and quantity might impact males and females differently.

A major standout from this analysis is the disparity between the SNP heritability estimates here and heritability estimates from the twin literature that are larger in magnitude. For example, Li, Cheng, Ma, and Swan (2003)'s meta-analysis reported heritability estimates for smoking initiation of .37 in females and .55 in males, and for persistence .46 in females and .59 in males. Analyses that specifically looked at age of smoking onset found heritability of about .60 (Morley et al., 2007). Even though GCTA confirmed the contribution of common SNPs to smoking onset and quantity, these estimates of the genetic contribution to smoking-

related traits are less than what family studies have previously found. However, an important consideration is that the genetic source of variance calculated by GCTA, is not the heritability, or proportion of trait variance due to additive genetic factors. It is simply the proportion of trait variance that can be attributed to measured SNPs that may *or may not* capture the entirety of additive genetic factors contributing to a trait.

Like Yang et al.'s (2010) initial study on human height, several other studies have sought to use GCTA in an attempt to rectify the discrepancy between heritability estimates from the family-based literature with the tiny genetic variance estimates obtained from SNPs in GWAS. For the most part, the additive genetic variance that can be obtained from the aggregate of all SNPs tends to be much greater than that obtained from single SNPs in GWAS (Gusev et al., 2013), but less than that estimated from family studies. For example, Lubke et al. (2012) examined smoking initiation (GCTA = .19/ TwinFam = .44), current smoking (.24/.79), fasting glucose (.22/ .53), and height (.42/.90). Further, Vinkhuyzen et al. used GCTA to get an estimate of the degree to which common SNPs explain variation in personality dimensions of neuroticism ($\sim$ 6%) and extraversion ($\sim$12%). Again, these estimates are much greater than what had been previously found in GWAS or candidate gene studies, but un-accounted for genetic variance, as estimated by family-based studies, remains.

There are a number of explanations for the lower estimates by GCTA compared to family study estimates of heritability. Wray and Maier give an extensive review of these possibilities (Wray & Maier, 2014). One explanation has

to do with the key difference between GCTA and family-style analyses. While family style analyses are supposedly picking up on total additive genetic variance due to the assumptions underlying the covariance structure between different relatives of differing genetic relatedness, the ability for GCTA to produce an accurate measure of genetic variance depends on the degree to which the common SNPs that were used in the study are correlated or in LD with the true causal variants. For many traits there has been suspicion that perhaps rare causal variants (with MAFs much lower than the marker), which would not be picked up as well by the genotyped common SNPs, contribute to genetic variance in the trait. In other words, the there may be imperfect tagging, either through LD or just missing particular parts of the genome, between the genotyped SNPs and the un-observed causal variants. Thus, Davies et al. (2011) in their study using this design for intelligence, describe GCTA as a way to get a lower bound estimate for true narrow-sense heritability.

A further reason for the discrepancy between genetic variance estimates from twin studies and analyses utilizing GCTA has to do with the possibility that twin studies may be incorrectly assessing narrow-sense heritability. This could happen for a variety of reasons that have to do with the assumptions and general ways twin models are specified. One possibility is that non-additive genetic factors, that are often un-modeled in classical twin designs, may be biasing shared environmental effects downward, and thus estimates of heritability will be increased. One example of this includes the potential for genetic interactions that are not considered in an additive model (Zuk, Hechter, Sunyaev, & Lander, 2012). Further any other effect that would bias heritability estimates in an upward fashion,

such as violations of the equal environment assumption, will contribute to this discrepancy. Additionally, when fitting twin models, dropping non-significant parameters, particularly those for shared-environment, when they are in fact contributing to twin co-variation might further bias heritability estimates upward and further widen the discrepancy between genetic variance derived from common SNPs using GCTA and that from family-based designs.

Additionally, twin studies examine the phenotypic correlation between individuals that are the same age while analyses in GCTA look at the phenotypic correlation between individuals that, depending on the sample used, could differ in age by several years. In the case of this sample, participants were born from the 1920s to the 1950s. Thus, comparisons between smoking behavior would be made between individuals of different ages. Thus, to the degree that age or cohort has influence on the genetic etiology of smoking behavior, we would be 'comparing apples to oranges' with respect to linking genetic similarity to phenotypic similarity as is done in GCTA. Wray and Maier (2014) discuss how disease heterogeneity, where different sets of genetic factors contribute to diseases that are classified as a single condition, may result in underestimates of SNP heritability. Correspondingly, if different sets of genetic factors are responsible for smoking phenotypes depending on a person's age or cohort, these lower estimates of SNP heritability for smoking onset and quantity would not be surprising. As family studies have hinted at potential differences in the genetic components related to smoking based on cohort (e.g. Boardman et al., 2010), we will examine this possibility in the next chapter.

CHAPTER V

GENOME-WIDE SNP DATA SUGGESTS GENETIC INFLUENCES ON CIGARETTE

SMOKING INITIATION DIFFER ACROSS BIRTH COHORTS

Background

Twin studies have highlighted the potential for heritability estimates to differ depending on the specific birth cohort examined (Heath et al., 1993; Boardman et al., 2010). Thus, we used a mixed linear effect model, implemented in Genome-wide Complex Trait Analysis (GCTA)(Yang et al., 2011), to understand whether the contribution of common genetic variants to smoking initiation differs as a function of birth cohort.

Similar to the referenced twin studies, differences between cohorts in the magnitude of SNP h^2 would point to a *quantitative* GE interaction. However, an advantage of using GCTA over family studies is that, because genetic similarity between all pairs of subjects are used to estimate SNP h^2, age differences between individuals in the sample allow estimation of the overlap between genetic effects responsible for smoking initiation across different ages, or *qualitative* GE interaction. In GCTA this can be done in two ways: (1) by estimating the genetic correlation between the two cohort groups in a bivariate analysis (Lee et al., 2012), and (2) by including a gene-by-cohort (qualitative) interaction term. Both approaches assess the degree to which the SNP effects associated with smoking initiation in one birth cohort are the same as those in another birth cohort.

We were interested in understanding both quantitative and qualitative gene-by-cohort interaction effects for smoking initiation during the 20th century, as the health consequences of smoking began to become apparent. We first demonstrate these approaches using a fake data simulation to ensure that the estimates of quantitative and qualitative GE interactions were unbiased and modeled appropriately. We then divided a large genome-wide dataset on smoking initiation into two birth cohorts to estimate the SNP h^2 in each cohort separately, and then tested both the genetic correlation across cohorts, and the qualitative gene-by-cohort interaction, in the combined sample.

<div align="center">Method</div>

Simulation of Quantitative and Qualitative GE Interaction in GCTA

We began our analysis by simulating polygenic quantitative or qualitative GE interactions. To approximate the underlying genetic structure and reduce computation time, we used real genotype data from chromosome 22 (3397 SNPs spanning 34 Mb) in the Atherosclerosis Risk in Communities study accessed from dbGaP. We divided the sample in half at random and called one group the 'old' cohort (n = 4140) and one the 'young' cohort (n = 4140). For both interaction models we simulated a phenotype with a prescribed SNP h^2 based on 150 randomly chosen causal variants which each contributed equal genetic variation to the phenotype. For quantitative GE interaction, the causal variants for each cohort were the same but the simulated phenotype had a SNP h^2 of .6 in the old cohort and .3 in the young cohort. To simulate a qualitative GE interaction, of the 150 causal variants, 50 were unique to the old cohort, 50 were unique to the young cohort, and

50 were shared by both. The SNP h^2 was set to .6 in each cohort. Thus, within each cohort, variance accounted for by SNPs unique to each cohort was .3 and variance related to causal SNPs shared by the two groups was also .3, implying a genetic correlation (SNP-r_g) of .50 and a r^2 (or h^2) of the gene-by-cohort interaction term of .3.

For both simulations we used GCTA to estimate the heritability for the cohorts combined and separately. We also tested an interaction between genotype and cohort and did a bivariate analysis between the two cohorts. We repeated this 100 times, each time creating new cohorts affected by new causal variants.

Real Data: ARIC and MESA

Genetic and phenotypic data obtained from dbGaP consisted of 10,162 participants from the Atherosclerosis Risk in Communities (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA). This is the same dataset used in Chapter IV, "Snps And Smoking: What Can The Aggregate Of Genome-Wide Snps Tell Us About Genetic Liability To Smoking Initiation And Quantity Smoked?" See the previous chapter for details and quality control procedures. Missing phenotypes reduced our sample to 8484 individuals.

Measures

We assessed smoking initiation in both datasets. In ARIC, initiation was assessed by interview with the question, "Have you ever smoked cigarettes?" In MESA, initiation was assessed by questionnaire with the question, 'Have you ever smoked at least 100 cigarettes in your lifetime?' Responses were a dichotomous 'yes' or 'no'. Differences in reported smoking behavior has not been found to differ

as a function assessment method (Kaplan, Hilton, Park-Tanjasiri, & Pérez-Stable,
2001). The proportion of smokers in MESA dataset (56%, n = 1942) was slightly
lower than in the ARIC dataset (61%, n = 8102). To correct for any dataset related
discrepancies, dataset was included as a covariate in all models.

Analyses

To have enough individuals within each cohort to ensure adequate power to
detect SNP-correlations or GE interactions, we split the sample based on the median
birth year of 1934, creating two cohort groups of between four and five thousand
individuals (Visscher et al., 2014). Individuals born before 1934 were called the 'old'
cohort and those born in or after 1934 were called the 'young' cohort.

To test for quantitative GE interaction, we estimated the SNP h^2 for the
young and old cohorts separately as well as for the combined sample (both old and
young cohorts). We estimated the genetic correlation between SNPs (SNP-r_g) in the
young and old cohorts using two methods. First, using the gene-environment
interaction option, we estimated a single additive genetic variance component that
is assumed/required to be equal in each cohort, an environmental variance
component, also assumed to be equal in each cohort, and a gene environment
variance component (V_{AxE}) that effectively parameterizes a genetic correlation. A
significant V_{AxE} component suggests that the genetic correlation is significantly
different from 1.0. Second, using the bivariate option, we estimated the additive
genetic and environmental variance for each cohort and the genetic correlation
between the cohorts, which did not require the genetic variances in the two cohorts
to be equal.

Additionally, due to the arbitrary nature of the median data split, we estimated the SNP h^2 for overlapping groups of individuals born in 8-year intervals. Further, we tested the interaction between gender and genotype both in a model that included just gender and, to confirm that gender interactions are not driving cohort interactions, in a model allowing for both cohort and gender interactions. Dataset and birth year (as a continuous instead of dichotomous variable) were also included as interacting factors with genotype in separate models. We included gender, dataset, and the first ten ancestry principal components (to control for ethnic stratification (Price et al., 2006)) as covariates in all analyses. Birth year was used as covariate in analyses in which it was not collinear with birth cohort. All analyses and quality control were conducted using PLINK v1.07(Purcell et al., 2007), PLINK v1.90b2b(Purcell & Chang, 2014), R version 3.0.1(Team 2013), and GCTA version 1.24.3(Yang et al., 2011b).

<div align="center">Results</div>

SNP Heritability and GE Interactions for Simulated Phenotypes

Results from the simulation of the quantitative GE interaction, are shown in Table 5.1. GCTA correctly estimated the SNP h^2 within each cohort, and when the cohorts were combined into a single sample, the SNP h^2 estimate was at an intermediate value between the values for the old and young cohort. Furthermore, the interaction term between genotype and cohort in the combined sample was null, and the bivariate analysis of the SNP correlation demonstrated that the causal SNPs in one cohort were identical to those in the other.

Results from the simulation of the qualitative GE interaction demonstrate

that the SNP h^2 for the old and young cohorts correctly estimated the genetic

variance from causal SNPs for each cohort (Table 5.1). When the two cohorts were

combined in a single sample, the estimated SNP h^2 was reduced to a level that was

lower than the SNP h^2 in either cohort, because the SNP h^2 estimate was depressed

by the genetic variance that was due to SNPs that differed between cohorts. The

interaction term between cohort and genotype correctly indicated that half of the

genetic variance was shared within and half between cohorts. In the bivariate

analysis, the SNP correlation also indicated that half the SNPs relevant to each group

were shared while the other half were unique to each cohort (SNP-$r_g \sim .50$).

Table 5.1

Simulation: GCTA estimates and standard errors

	Quant. Simulation (Mean SNP h^2)		Qual. Simulation (Mean SNP h^2)	
	n	Est. (SE)	n	Est. (SE)
Old Cohort	4140	.60 (.04)	4140	.61(.02)
Young Cohort	4140	.30 (.05)	4140	.60 (.02)
Old and Young	8280	.44 (.03)	8280	.46 (.02)
Interaction	8280		8280	
V(G)/V(P)		.42 (.03)		.32 (.02)
V(GxCohort)/V(P)		.04 (.03)		.28 (.02)
Bivariate	8280		8280	
V(G)/V(P) Old		.60 (.04)		.61 (.02)
V(G)/V(P) Young		.31 (.04)		.60 (.02)
r_g SNP Young-Old		.96 (.07)		.53 (.03)

SNP Heritability and GE Interactions for Smoking Initiation

Sixty percent of the old cohort had initiated smoking (48% prevalence in females and 72% in males), while 59% had initiated in the young cohort (52% prevalence in females and 67% in males). There was not a significant interaction between genotype and dataset (V(GxDataset)/V(P) = 0; SE = .13; p = .5), indicating that the same SNPs predicted smoking initiation in the two datasets, which provided justification for combining them.

Results from the main GCTA analyses are shown in Table 5.2. For individuals born before 1934, SNP h^2 for smoking initiation was not significant (SNP h^2 = .15; SE = .13; p = .12), but there was a significant SNP h^2 for individuals born in or after 1934 (SNP h^2 = .31; SE = .11; p = .002). There was no evidence that these two SNP h^2 estimates differed (that there was a quantitative gene-by-cohort interaction; Z = -1.37, p = .17). Figure 5.1 illustrates the SNP h^2 estimates for overlapping sets of individuals born in cohorts of 8 years, suggesting that the strongest genetic contribution to smoking initiation occurred for those born in the 1930s.

Table 5.2

ARIC and MESA data: GCTA estimates and standard errors

	n	SNP h^2
Old Cohort	3912	.15 (.13)
Young Cohort	4572	.31 (.11)
Old and Young	8484	.13 (.06)
Interaction	8484	
V(G)/V(P)		.003 (.08)
V(GxCohort)/V(P)		.23 (.12)
Bivariate	8484	
V(G)/V(P) Old		.15 (.13)
V(G)/V(P) Young		.31 (.11)
r_g SNP Young-Old		-.09 (.40)
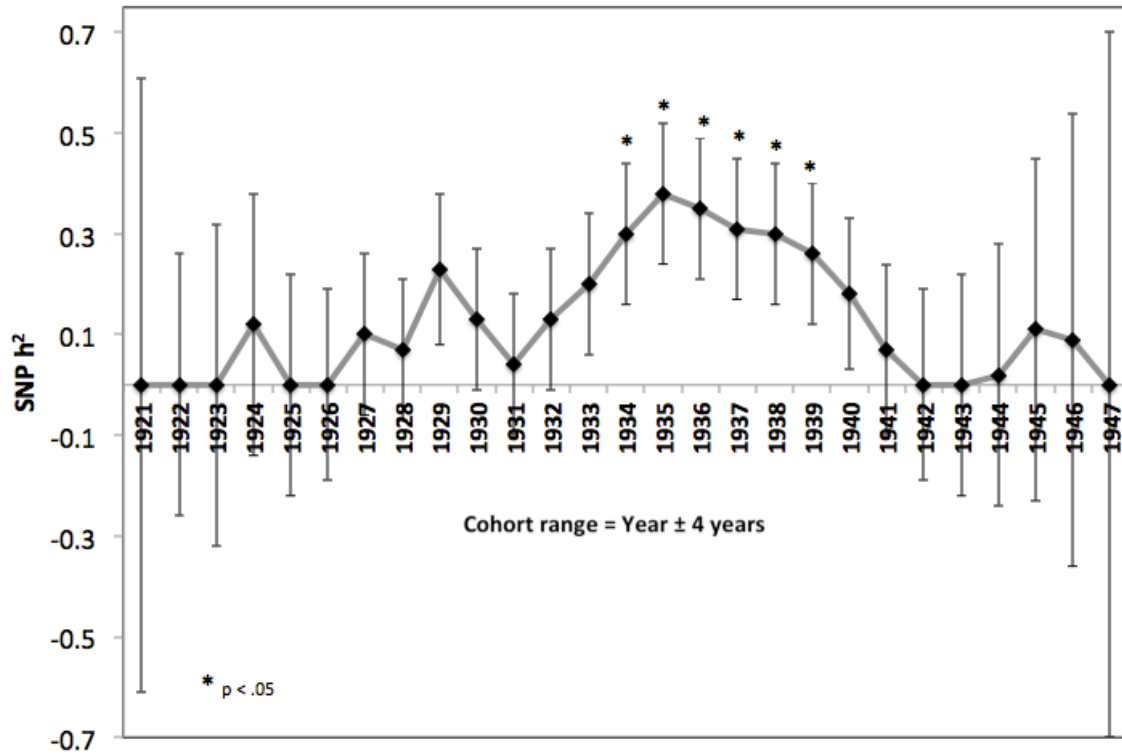
Figure 5.1. Estimated SNP heritability for overlapping sets of individuals (points

non-independent) born in cohorts of 8 years. We used Genome-wide Complex Trait

Analysis (GCTA) to estimate SNP heritability (SNP h^2) for individuals born in

overlapping 8-year intervals (± 4 years from the year shown). Significant estimates

indicate significant SNP heritability estimates for individuals born within the 8-year

interval. As displayed, the greatest genetic influence indicated by significant SNP h^2

estimates was found for individuals born in the 1930s. Error bars represent

standard errors.

In the combined old and young cohort groups, there was a significant effect of common SNPs (SNP h^2 = .13; SE = .06; p = .02) on smoking initiation, which was lower in magnitude than in either the old or young cohort when analyzed separately (see Table 5.1), suggestive of a qualitative birth cohort-by-SNP interaction. Consistent with this expectation, there was a significant interaction between the effect of common SNPs and cohort (V(GxCohort)/V(P) = .23; SE = .12; p = .02) for smoking initiation. To ensure this result was not confounded by gender (Keller 2014), we included a gender-by-gene interaction in this model; the cohort interaction remained significant (V(GxCohort)/V(P) = .22; SE = .12; p = .03) while the gender interaction was not (V(GxGender)/V(P) = .09; SE = .11; p = .22). In the bivariate analysis, the estimate of SNP-r_g was low (-.09; SE = .40), suggesting that SNP effects associated with smoking initiation differ across birth cohorts. This estimate is significantly different from 1.0 using a Z test (Z = -2.74, p = .003) but not by a likelihood ratio test (χ^2_1 = 1.49, p = .11). However, when we replaced the dichotomous (young/old) variable with birth year as a continuous variable, evidence for the cohort-by-gene interaction disappeared (V(GxYear)/Vp = 0; SE = .29, p = .5).

Discussion

We used a mixed linear effect method to understand how genetic factors responsible for smoking initiation tagged by genome-wide SNPs differ in terms of magnitude and composition for two birth cohorts. Genetic influences as indexed by the conglomerate of genome-wide SNPs contributed to smoking initiation for those born in or after 1934, but were lower and non-significant for those born before

1934. The heritability estimate in the combined cohort analysis was less than that for either cohort individually, suggestive of a low genetic correlation between smoking initiation in the two cohorts, and a SNP correlation significantly less than 1 was found in the bivariate analysis. Additionally, a significant genotype-by-cohort qualitative interaction term was found in the combined cohort analysis that further supported the hypothesis that genetic effects responsible for smoking initiation differ across birth cohorts.

An advantage of using the qualitative interaction term in addition to the bivariate analysis is that we could include a continuous interacting factor instead of dichotomizing birth year into two cohorts. However, when we treated cohort as a continuous variable (birth year), evidence for the cohort-by-gene interaction disappeared. We are uncertain why this occurred, but one possibility is nonlinearity in the birth year-by-gene interaction effect (See Figure 5.1). Additionally, the qualitative interaction analysis requires homogeneity of genetic and environmental variance across the groups being compared, and violations of this could lead to false positive results. The bivariate analysis relaxes this assumption. Yet, both analyses suggested that the genetic components or effects related to smoking initiation might be different depending on cohort.

The arbitrary median split prompted us to examine differences in estimates of SNP h^2 for 8-year increments across our entire sample. Our results mirrored findings from Boardman et al. (2010) who found genetic influences for individuals born in the 1930s. Low sample sizes in the tail birth years of our sample hindered our ability to get accurate SNP h^2 estimates for cohorts defined by specific events,

such as the 1964 Surgeon General's Report, that are likely to have influenced smoking behavior. Larger samples that span greater age ranges could test the same GE interactions investigated here across potentially more informative cohorts.

Gender may also have an impact on the quantitative or qualitative differences in heritability across birth cohort observed here, but splitting the sample by both gender and cohort would increase the imprecision of SNP h^2 estimates for each group. Nevertheless, it is possible that our findings could be the result of genetic factors that are differentially related to smoking initiation for males versus females. However, given that the cohort interaction remained significant even when accounting for potential gender interaction, it is unlikely that gender differences could be entirely driving our finding of genetic differences by cohort.

This analysis focused on smoking initiation and did not examine these effects for other smoking phenotypes, such as nicotine dependence, which might show different patterns of heritability and environmental moderation. Further, as seen in the previous chapter, these estimates of SNP h^2 are lower than heritability estimates from twin/family studies (Sullivan & Kendler, 1999; Li et al., 2003) primarily because only effects of (mostly common) genetic variants tagged by SNPs are included in the estimate of SNP h^2, but also because SNP h^2 estimates are not inflated by common environmental or non-additive genetic effects (Keller & Coventry, 2005). Finally, motivated by our need to maximize sample size, we used a $\hat{\pi}$ threshold (< .05) that was more relaxed than often used (<.025) (Yang et al., 2010). As this threshold is increased, it is increasingly possible that shared environmental factors (shared between very extended families) to be confounded

with genetic factors, which could lead to inflated estimates of SNP h^2. Nevertheless, we find this possibility unlikely given the very distant relatedness implied by the low threshold for genetic similarity used in this analysis.

We illustrated GCTA's ability to detect differences in not only the magnitude of genetic effects but in the actual genetic factors that affect the same trait at two time periods. Importantly, there might be different genetic effects influencing smoking initiation for those born either before or after the mid-1930s. To better understand the dynamic of the social environment and any gender effects that could differentially impact the each of these cohorts, obtaining larger genome-wide datasets with an expanded range of birth years is a crucial next step. Such inquiry will allow a clearer understanding of how the genetic underpinnings of smoking exist in synergy with the environmental and social context.

CHAPTER VI

OVERALL DISCUSSION, LIMITATIONS AND IMPLICATIONS

These four studies explore some of the primary contributors to the etiology

of cigarette smoking. The influences of genes, peers, and birth cohort have all been

previously described in terms of their unique relationship to various definitions of

cigarette smoking behavior. Here, however, we further clarified how these factors

may synergistically influence choices to smoke.

Studies 1 and 2

In Study 1, we described the various mechanisms that explain homogeneity

in smoking behavior among peers. Specifically, we addressed the possibility that

shared genes may put individuals into friendship groups that may then provide

additional peer influence. Additionally, we assessed whether this mechanism is

limited to the initiation of smoking or whether such processes are still relevant for

heavier use of cigarettes.

Our findings demonstrated that homophily, or selecting friends that are

genetically similar to oneself in terms of smoking behavior, could partially explain

why individuals resemble their peers with respect to smoking behavior, especially

with respect to smoking initiation. We also described a number of alternatives that

may explain smoking homogeneity in peer groups, but our specific pattern of

correlations points to the presence of genetically based homophily.

However, we were limited in our ability to evaluate the presence of active gene environment correlation (active rGE), which would require that we also have information on the degree to which the selected peer group provided additional influence on smoking behavior. Ideally, longitudinal data could be used to assess both the effects of peer selection and the possibility of influence processes, especially if information was known about the smoking behavior of the friends. Additional limitations further discussed in the relevant chapter included: the young age of participants, relevance of the questions related to smoking behavior (especially for high school students), and the ability to extend these results to more modern cohorts of adolescents.

Study 2 uses a more recent dataset to expand upon Study 1 for two smoking phenotypes: smoking experimentation and regular use of cigarettes. Again, we sought to determine the degree to which shared friends influences similarity of twin smoking. Our findings demonstrated that shared peers among twins contribute to their similarity of smoking for both smoking experimentation and more regular use of cigarettes. However, we only see a replication of the results from Study 1, an effect for DZ but not MZ twins, for regular smoking at Wave 2 (W2). This result could demonstrate that choosing friends like oneself with respect to genetic propensity for smoking may only be a viable explanation for peer smoking homogeneity for older individuals that are using cigarettes more regularly. Rather, we remain cautious in interpreting this result due to the reduced sample size in W2 and the high initial correlation for MZ twins. Other analyses have taken advantage of the longitudinal aspects of this dataset to demonstrate the salience of peer

selection, but have found limited evidence for additional influence (Hoffman et al., 2007).

In conclusion, both Study 1 and 2 established that selecting peers like oneself contributes to why peers resemble each other in terms of smoking behavior. Study 1 specifically pointed to the possibility that this selection is based upon genetic similarity in domains related to smoking initiation. Study 2 demonstrated that friend selection is also an important precursor to both smoking experimentation and regular use. However, we had limited evidence with regard to the additional influence that would be needed to conclude the presence of active rGE. Longitudinal data that assesses individual and their friends with regard to smoking over more expanded time periods would be key in trying to assess the impact of active rGE.

Though we remain unsure about the potential for active rGE with respect to smoking behavior, we were able to evidence homophily (selection) as a driving mechanism in why individuals resemble their peers for both smoking experimentation/initiation and more regular smoking habits. Thus, in an effort to deter adolescents from smoking, it may be a more effective strategy, instead of campaigns that largely target peer pressure and influence, to begin intervention with the encouragement of the formation of friends that that exhibit similar healthy habits.

Studies 3 and 4

Studies 3 and 4 switched from the use of family data to whole-genome SNP data. These analyses occurred in the context of what is often termed 'missing heritability' or the apparent inability to recover from single SNPs through genome-

wide association studies (GWAS) the genetic variance estimated from family studies for behavioral traits (Maher 2008; Manolio et al., 2009). However, using the same SNP data originally collected for traditional GWAS, recent methods have been developed to possibly circumvent some of the limitations related to GWAS for behavioral traits (Visscher, Brown, McCarthy, & Yang, 2012). Some examples include polygenic risk scores (Purcell et al., 2009), pathway analyses (Wang, Li, & Hakonarson, 2010), and, as we will focus on here, analyzing the simultaneous effect of all SNPs (Yang et al., 2010). Thus, Studies 3 and 4 used mixed linear effects models to estimate the genetic variance and genetic correlation due to the conglomerate of all genome-wide SNPs using a program called Genome-wide Complex Trait Analysis (GCTA) (Lee et al., 2012b; Yang et al., 2011b). Both of these analyses used genome-wide SNP data from dbGaP.

In Study 3, we focused on two smoking phenotypes, smoking onset and quantity. Genetic factors as indexed by the conglomerate of genome-wide SNPs were able to account for variance in both the age of smoking onset and quantity of cigarettes smoked. Further, it appears that there may be unique genetic factors related to the quantity of cigarettes smoked. However, when analysis was divided into male and female samples, we were unable to detect significant genetic influences for each phenotype. Further, there do not appear to be unique genetic factors that influence each gender's propensity to start smoking at a particular age or smoke more cigarettes per day.

While these results remained ambiguous with respect to effects for each gender, we did demonstrate that genome-wide SNPs in conglomerate can account

for variance in smoking phenotypes. In relation to The Tobbacco and Genetics

Consortium (2010) that found a number of individual SNPs that contributed to both

smoking initiation and cigarettes smoked per day, using the conglomerate of all

SNPs, we were able to account for 8-12% of the phenotypic variance for these

smoking phenotypes, more than could be accounted for by any single SNP. However,

again in comparison to the Consortium's results, using the conglomerate of SNPs

does not give any information with regard to *which* specific SNPs contribute to the

trait, making it impossible to connect this source of genetic variance to any genes or

biological pathways. Thus, it still remains important to consider a wide range of

approaches to fully understand the genetic etiology and underlying biology of

smoking. An additional consideration is that the 'SNP heritability' estimates

achieved here are considerably lower than reports from family studies. Briefly, this

may occur for a number of reasons including inability of the SNPs to pick up on the

true causal variants, overestimates of heritability in family studies, and genetic

heterogeneity. For a detailed discussion of these and other possibilities, see

Discussion in Chapter IV.

Beyond the ability to quantify the genetic contribution to smoking onset and

quantity was the goal of estimating the shared genetic variance between these two

traits. Results, demonstrating a low to moderate correlation between the SNPs

relevant to onset and quantity, replicate findings in the literature that evidenced

some unique genetic contribution to more advanced stages of smoking beyond

smoking initiation. This finding has particular implications in understanding

individual differences in smoking patterns, as it appears that individuals with a

genetic predisposition to initiate smoking may not have that same predisposition to smoke in quantity, and vice versa, though the affinity to smoke in quantity could be averted through non-initiation. As such, this finding is important methodologically; as it illustrates that there may be separate biological pathways related to the specific smoking stage in question.

The possibility that the lower 'SNP heritability' for smoking onset and quantity than reported in family studies could be due to differences in the ages of the individuals being compared in GCTA prompted us to examine cohort differences in both the magnitude and composition/effect of the genetic factors related to smoking initiation. Family studies have previously examined differences in the magnitude of these effects, though the results are mixed depending on the specific cohort and gender of the sample (Boardman et al., 2010; Heath et al., 1993; Kendler et al., 2000; Vink & Boomsma, 2011). Differences in the composition of these genetic factors are not feasibly obtained in family studies due to the need to compare individuals of different ages. However, GCTA, which compares unrelated individuals, gave us the unique ability to ask specifically whether the genetic components related to smoking initiation are shared by different cohorts.

In this analysis we divided individuals in our sample into 'young' and 'old' cohort groups based upon the median birth year of 1934. The conglomerate of genome-wide SNPs significantly contributed to variance in smoking initiation for individuals born in or after 1934, but not for individuals born prior. While this difference in the magnitude of heritability estimates was not significant (quantitative GE interaction), we did find evidence that there may be differences in

the composition or effect of genetic factors responsible for smoking initiation for those born before versus after 1934. Further, the greatest genetic effect here was for individuals born in the 1930s.

A major limitation here was that the sample did not include many individuals that would have started smoking after large milestones, such as the 1964 Surgeon General's Report that linked smoking to lung cancer. Thus, we do not have any specific hypotheses regarding why we see this difference in the genetic factors for the two cohorts. In fact, the sample sizes in the tail ends of our range of birth years were quite small, so we were not able to get good estimates of how the heritability of smoking initiation may have changed over time (See Figure 5.1). Importantly, however, we draw attention to the potential for not only the magnitude of heritability for smoking initiation to change over time but for the actual genetic factors to change according to the environmental context as well. Further, we demonstrate the use of a recent method to answer a question, whether the genetic factors related to smoking initiation are shared by individuals born during different time periods, that has previously been difficult to assess using family-based methods.

<div align="center">Overall Conclusion and Next Steps</div>

This dissertation examines the interplay of a number of factors that have been associated with cigarette smoking: genes, peers, stage of smoking, gender, and cohort. In Studies 1 and 2, we discovered that peers resemble each other in terms of smoking behavior due to homophily or selecting friends that are like oneself. We also have evidence that this occurs not only for initial experimentation with

cigarettes, but for more regular smoking as well. However, the role of active rGE in peer smoking homogeneity is unclear, as we do not know whether the chosen peer environment exerts additional influence on smoking behavior. This question could be resolved with longitudinal data spanning an adequate time range that assesses both the smoking behavior of both individual and their friends.

In studies 3 and 4 we used the conglomerate of SNPs from whole-genome data to carry out analyses using GCTA (Yang et al., 2011b). In Study 3, we learned that common SNPs do explain variance in smoking onset and quantity smoked. Additionally, there are both shared and unique SNPs related to smoking onset versus quantity. Given larger sample sizes, this analysis should be extended to look at the potential for gender effects as well as additional smoking phenotypes. Study 4 examined cohort and smoking. Not only were we able to discern whether heritability estimates for smoking initiation differ as a function of year born, but we were also able to examine whether the actual genetic factors related to smoking initiation are shared by individuals born during different time periods. While the magnitude of heritability did not differ by cohort, the data suggest there were different genetic factors contributing to the heritability of each. Larger datasets that span a wider range of birth years would be instrumental in allowing us to assess specific hypotheses related to changes in the genetic components, potential for these patterns to vary by gender, and whether these patterns are seen for smoking phenotypes besides initiation.

In summary, we have taken known factors related to cigarette smoking and, using new or underutilized methods hope to understand these factors in a new light.

These approaches not only allow us to understand the direct relationship between genes, peers, cohort and smoking, but allow for insight into the complex ways these factors may dually influence choices to smoke.

Additionally, many of these methods and insights are applicable to a host of traits that ensue a number of additional queries. For example, is peer selection (and its potential to contribute to active rGE) relevant for other traits that we see shared by peer groups? Further, given our finding that there are unique genetic factors related to smoking in quantity, to what degree are these unique genetic factors related to addiction in general? The ability to use GCTA to estimate shared genetic variance using unrelated individuals creates even more possibilities to understand the role of genetics in connecting particular phenotypes, especially in cases where the use of family data would be infeasible for the particular question. For example, psychiatric conditions are often accompanied by substance use; to what degree is this phenotypic correlation the result of shared genetic etiology? Thus, the methodology and findings presented here do more than elucidate some of the factors relevant to cigarette smoking, but open up a series of additional ways to understand the complex intertwine of the genetic and environmental etiology of smoking, substance use, addiction, and behavior.

CHAPTER VII

REFERENCES

Alexander, C., Piazza, M., Mekos, D., & Valente, T. (2001). Peers, schools, and adolescent cigarette smoking. *J Adolesc Health*, *29*(1), 22-30.

Amos, C. I., Spitz, M. R., & Cinciripini, P. (2010). Chipping away at the genetics of smoking behavior. *Nat Genet*, *42*(5), 366-8. doi:10.1038/ng0510-366

Arnett, J. J. (2007). The myth of peer influence in adolescent smoking initiation. *Health Education & Behavior*, *34*(4), 594-607.

Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Roux, A. V. D., Folsom, A. R., . . . Liu, K. (2002). Multi-ethnic study of atherosclerosis: Objectives and design. *American Journal of Epidemiology*, *156*(9), 871-881.

Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *J Health Soc Behav*, *51*(1), 108-123.

Boardman, J. D., Saint Onge, J. M., Haberstick, B. C., Timberlake, D. S., & Hewitt, J. K. (2008). Do schools moderate the genetic determinants of smoking? *Behavior Genetics*, *38*(3), 234-46. doi:10.1007/s10519-008-9197-0

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Boomsma, D. I., Koopmans, J. R., Van Doornen, L. J., & Orlebeke, J. F. (1994). Genetic and social influences on starting to smoke: A study of dutch adolescent twins and their parents. *Addiction (Abingdon, England)*, *89*(2), 219-26.

Carey, G. (1988). Inference about genetic correlations. *Behavior Genetics*, *18*(3), 329-338.

Centers for Disease Control and Prevention (CDC). (2012). Current cigarette smoking among adults - united states, 2011. *MMWR Morb Mortal Wkly Rep*, *61*(44), 889-94.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., . . . Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry*, *16*(10), 996-1005. doi:10.1038/mp.2011.85

Degenhardt, L., Chiu, W. T., Sampson, N., Kessler, R. C., & Anthony, J. C. (2011). Epidemiological patterns of extra-medical drug use in the united states: Evidence from the national comorbidity survey replication, 2001-2003. *Drug Alcohol Depend*, *90*(2-3), 1-14. doi:10.1016/j.drugalcdep.2007.03.007

Dishion, T. J., & Owen, L. D. (2002). A longitudinal analysis of friendships and substance use: Bidirectional influence from adolescence to adulthood. *Dev Psychol*, *38*(4), 480-91.

Fisher, R. A. (1958). Lung Cancer And Cigarettes. *Nature*, *182*(4628), 108.

Friberg, L., Kaij, L., Dencker, S. J., & Jonsson, E. (1959). Smoking habits of monozygotic and dizygotic twins. *Br Med J*, *1*(5129), 1090-2.

Go, M. -H., Green, H. D., Kennedy, D. P., Pollard, M., & Tucker, J. S. (2010). Peer influence and selection effects on adolescent smoking. *Drug Alcohol Depend*, *109*(1-3), 239-42. doi:10.1016/j.drugalcdep.2009.12.017

Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B. J., Diogo, D., Stahl, E. A., . . . Price, A. L. (2013). Quantifying missing heritability at known GWAS loci. *PLoS Genetics*, *9*(12), e1003993. doi:10.1371/journal.pgen.1003993

Hall, J. A., & Valente, T. W. (2007). Adolescent smoking networks: The effects of influence and selection on future smoking. *Addict Behav*, *32*(12), 3054-9. doi:10.1016/j.addbeh.2007.04.008

Hamilton, A. S., Lessov-Schlaggar, C. N., Cockburn, M. G., Unger, J. B., Cozen, W., & Mack, T. M. (2006). Gender differences in determinants of smoking initiation and persistence in california twins. *Cancer Epidemiol Biomarkers Prev*, *15*(6), 1189-97. doi:10.1158/1055-9965.EPI-05-0675

Hardie, T. L., Moss, H. B., & Lynch, K. G. (2006). Genetic correlations between smoking initiation and smoking behaviors in a twin sample. *Addictive Behaviors*, *31*(11), 2030-2037. doi:10.1016/j.addbeh.2006.02.010

Harris, K. M. (2011). Design features of add health. *Carolina: University of North Carolina at Chapel Hill.*

Hartz, S. M., Short, S. E., Saccone, N. L., Culverhouse, R., Chen, L., Schwantes-An, T. -H., . . . Bierut, L. J. (2012). Increased genetic vulnerability to smoking at CHRNA5 in early-onset smokers. *Archives of General Psychiatry*, *69*(8), 854-860. doi:10.1001/archgenpsychiatry.2012.124

Heath, A. C., Cates, R., Martin, N. G., Meyer, J., Hewitt, J. K., Neale, M. C., & Eaves, L. J. (1993). Genetic contribution to risk of smoking initiation: Comparisons across birth cohorts and across cultures. *Journal of Substance Abuse*, *5*(3), 221-246.

Heath, A. C., Martin, N. G., Lynskey, M. T., Todorov, A. A., & Madden, P. A. (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*(02), 113-124.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K. O. (1991). The
fagerström test for nicotine dependence: A revision of the fagerström
tolerance questionnaire. *British Journal of Addiction*, *86*(9), 1119-27.

Hoffman, B. R., Monge, P. R., Chou, C. -P., & Valente, T. W. (2007). Perceived peer
influence and peer selection on adolescent smoking. *Addictive Behaviors*,
32(8), 1546-1554. doi:10.1016/j.addbeh.2006.11.016

Hoffman, B. R., Sussman, S., Unger, J. B., & Valente, T. W. (2006). Peer influences on
adolescent cigarette smoking: A theoretical review of the literature. *Subst Use
Misuse*, *41*(1), 103-55. doi:10.1080/10826080500368892

Holliday, J. C., Rothwell, H. A., & Moore, L. A. R. (2010). The relative importance of
different measures of peer smoking on adolescent smoking behavior: Cross-
sectional and longitudinal analyses of a large british cohort. *J Adolesc Health*,
47(1), 58-66. doi:10.1016/j.jadohealth.2009.12.020

Hughes, J. R. (1986). Genetics of smoking: A brief review. *Behavior Therapy*, *17*(4),
335-345.

Investigators, A. (1989). THE ATHEROSCLEROSIS RISK IN COMMUNITIES (ARIC)
STUDY: DESIGN AND OBJECTIVES. *American Journal of Epidemiology*, *129*(4),
687-702.

Kaplan, C. P., Hilton, J. F., Park-Tanjasiri, S., & Pérez-Stable, E. J. (2001). The effect of data collection mode on smoking attitudes and behavior in young african american and latina women. Face-to-face interview versus self-administered questionnaires. *Eval Rev*, *25*(4), 454-73.

Keller, M. C. (2014). Gene × environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, *75*(1), 18-24. doi:10.1016/j.biopsych.2013.09.006

Keller, M. C., & Coventry, W. L. (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *8*(3), 201-13. doi:10.1375/1832427054253068

Kendler, K. S., & Gardner, C. O. (1998). Twin studies of adult psychiatric and substance dependence disorders: Are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence? *Psychological Medicine*, *28*(3), 625-33.

Kendler, K. S., Thornton, L. M., & Pedersen, N. L. (2000). Tobacco consumption in swedish twins reared apart and reared together. *Arch. Gen. Psychiatry*, *57*(9), 886-892.

Kobus, K. (2003). Peers and adolescent smoking. *Addiction (Abingdon, England)*, *98 Suppl 1*, 37-55.

Koopmans, J. R., Slutske, W. S., Baal, G. C. M. V., & Boomsma, D. I. (1999). The influence of religion on alcohol use initiation: Evidence for genotype X

environment interaction. *Behavior Genetics*, *29*(6), 445-453.

doi:10.1023/A:1021679005623

Koopmans, J. R., Slutske, W. S., Heath, A. C., NEALE, M. C., & Boomsma, D. I. (1999).

The genetics of smoking initiation and quantity smoked in dutch adolescent

and young adult twins. *Behavior Genetics*, *29*(6), 383-393.

doi:10.1023/A:1021618719735

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of

pleiotropy between complex diseases using single-nucleotide polymorphism-

derived genomic relationships and restricted maximum likelihood.

Bioinformatics, *28*(19), 2540-2542. doi:10.1093/bioinformatics/bts474

Li, M. D., Cheng, R., Ma, J. Z., & Swan, G. E. (2003). A meta-analysis of estimated

genetic and environmental effects on smoking behavior in male and female

adult twins. *Addiction (Abingdon, England)*, *98*(1), 23-31. doi:10.1046/j.1360-

0443.2003.00295.x

Loehlin, J. C. (2010). Is there an active gene-environment correlation in adolescent

drinking behavior? *Behavior Genetics*, *40*(4), 447-51. doi:10.1007/s10519-

010-9347-z

Loehlin, J. C., & Nichols, R. C. (1976). Heredity, environment, & personality: A

study of 850 sets of twins. *Heredity, Environment, & Personality: A Study

of 850 Sets of Twins.*

Lopez, A. D., Collishaw, N. E., & Piha, T. (1994). A descriptive model of the cigarette

epidemic in developed countries. *Tobacco Control*, *3*(3), 242-247.

doi:10.1136/tc.3.3.242

Lubke, G. H., Hottenga, J. J., Walters, R., Laurin, C., de Geus, E. J., Willemsen, G., . . . Boomsma, D. I. (2012). Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biological Psychiatry*, *72*(8), 707-709. doi:10.1016/j.biopsych.2012.03.011

Madden, P. A., Heath, A. C., Pedersen, N. L., Kaprio, J., Koskenvuo, M. J., & Martin, N. G. (1999). The genetics of smoking persistence in men and women: A multicultural study. *Behavior Genetics*, *29*(6), 423-31.

Maes, H. H., Sullivan, P. F., Bulik, C. M., Neale, M. C., Prescott, C. A., Eaves, L. J., & Kendler, K. S. (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine*, *34*(7), 1251-61.

Maes H. H., Neale M. C., Kendler K. S., Martin N. G., Heath A. C., Eaves L.J. (2006). Genetic and cultural transmission of smoking initiation: An extended twin kinship model. *Behav Genet, 36*(6):795-808.

Maher, B. (2008). The case of the missing heritability. *Nature*, *456*(7218), 18-21.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Chakravarti, A. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415-444.

Mercken, L., Snijders, T. A. B., Steglich, C., Vertiainen, E., & de Vries, H. (2010). Smoking-based selection and influence in gender-segregated friendship

networks: A social network analysis of adolescent smoking. *Addiction (Abingdon, England)*, *105*(7), 1280-9. doi:10.1111/j.1360-0443.2010.02930.x

Morley, K. I., Lynskey, M. T., Madden, P. A. F., Treloar, S. A., Heath, A. C., & Martin, N. G. (2007). Exploring the inter-relationship of smoking age-at-onset, cigarette consumption and smoking persistence: Genes or environment? *Psychological Medicine*, *37*(9), 1357-67. doi:10.1017/S0033291707000748

Munafò, M. R., & Johnstone, E. C. (2008). Genes and cigarette smoking. *Addiction (Abingdon, England)*, *103*(6), 893-904. doi:10.1111/j.1360-0443.2007.02071.x

National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. (2014). *The health consequences of smoking—50 years of progress: A report of the surgeon general.* Atlanta (GA): Centers for Disease Control and Prevention (US).

Pacheco, J. (2011). Trends—public opinion on smoking and anti-smoking policies. *Public Opin Q*, *75*(3), 576-592. doi:10.1093/poq/nfr031

Pollard, M. S., Tucker, J. S., Green, H. D., Kennedy, D., & Go, M. -H. (2010). Friendship networks and trajectories of adolescent tobacco use. *Addictive Behaviors*, *35*(7), 678-685. doi:10.1016/j.addbeh.2010.02.013

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, *38*(8), 904-9. doi:10.1038/ng1847

Purcell, S., & Chang, C. (2014). *PLINK 1.9*. Available from: https://www.cog-genomics.org/plink2.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . .

    Sham, P. C. (2007). PLINK: A tool set for whole-genome association and

    population-based linkage analyses. *American Journal of Human Genetics*,

    81(3), 559-75. doi:10.1086/519795

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F.,

    & Sklar, P. (2009). Common polygenic variation contributes to risk of

    schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748-52.

    doi:10.1038/nature08185

Simons-Morton, B. G., & Farhat, T. (2010). Recent findings on peer group influences

    on adolescent smoking. *J Prim Prev*, *31*(4), 191-208. doi:10.1007/s10935-

    010-0220-x

Sullivan, P. F., & Kendler, K. S. (1999). The genetic epidemiology of smoking. *Nicotine*

    Tob Res, *1*(Suppl 2), S51-S57. doi:10.1080/14622299050011811

Team, R. C. (2013). *R: A language and environment for statistical computing.* Vienna,

    Austria: R Foundation for Statistical Computing.

Timberlake, D. S., Rhee, S. H., Haberstick, B. C., Hopfer, C., Ehringer, M., Lessem, J. M., .

    . . Hewitt, J. K. (2006). The moderating effects of religiosity on the genetic and

    environmental determinants of smoking initiation. *Nicotine Tob Res*, *8*(1),

    123-33. doi:10.1080/14622200500432054

Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify

    multiple loci associated with smoking behavior. *Nat Genet*, *42*(5), 441-7.

    doi:10.1038/ng.571

Todd, G. F., & Mason, J. I. (1959). Concordance of smoking habits in monozygotic and dizygotic twins. *Heredity*, *13*(1959.417).

Van Reek, J., & Drop, M. J. (1986). Cigarette smoking in the USA: Sociocultural influences. *Revue D'épidémiologie Et De Santé Publique*, *34*(3), 168-73.

Vink, J. M., & Boomsma, D. I. (2011). Interplay between heritability of smoking and environmental conditions? A comparison of two birth cohorts. *BMC Public Health*, *11*(1), 316. doi:10.1186/1471-2458-11-316

Vink, J. M., Willemsen, G., & Boomsma, D. I. (2005). Heritability of smoking initiation and nicotine dependence. *Behavior Genetics*, *35*(4), 397-406. doi:10.1007/s10519-004-1327-8

Vink, J. M., Willemsen, G., Engels, R. C. M. E., & Boomsma, D. I. (2003). Smoking status of parents, siblings and friends: Predictors of regular smoking? Findings from a longitudinal twin-family study. *Twin Res*, *6*(3), 209-17. doi:10.1375/136905203765693861

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*(1), 7-24. doi:10.1016/j.ajhg.2011.11.029

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G. -B., Lee, S. H., Wray, N. R., . . . Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genetics*, *10*(4), e1004269. doi:10.1371/journal.pgen.1004269

Wang, K., Li, M., & Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. *Nature Reviews. Genetics*, *11*(12), 843-54. doi:10.1038/nrg2884

White, V. M., Hopper, J. L., Wearing, A. J., & Hill, D. J. (2003). The role of genes in tobacco smoking during adolescence and young adulthood: A multivariate behaviour genetic investigation. *Addiction (Abingdon, England)*, *98*(8), 1087-1100.

Wray, N. R., & Maier, R. (2014). Genetic basis of complex genetic disease: The contribution of disease heterogeneity to missing heritability. *Current Epidemiology Reports*, *1*(4), 220-227. doi:10.1007/s40471-014-0023-3

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., . . . Visscher, P. M. (2010). Common snps explain a large proportion of the heritability for human height. *Nat Genet*, *42*(7), 565-9. doi:10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011a). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76-82. doi:10.1016/j.ajhg.2010.11.011

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011b). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76-82. doi:10.1016/j.ajhg.2010.11.011

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1193-8. doi:10.1073/pnas.1119675109