2012

On the genetic and environmental associations between body composition, depression symptoms and smoking behavior.

Roseann Peterson
Virginia Commonwealth University

On the genetic and environmental associations between body composition, depression symptoms and smoking behavior.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By

Roseann E. Peterson
Bachelor of Arts, Biology, Minor Chemistry, University of Minnesota, 2004
Bachelor of Arts, Psychology, University of Minnesota, 2005

Advisor: Hermine H. Maes, Ph.D.
Associate Professor
Human and Molecular Genetics
Psychiatry
Massey Cancer Center
Virginia Institute for Psychiatric and Behavioral Genetics

Doctoral Committee:
Lindon J. Eaves, Ph.D., D.Sc.
Kenneth S. Kendler, M.D.
Brion S. Maher, Ph.D.
Suzanne E. Mazzeo, Ph.D.
Michael C. Neale, Ph.D.
Bradley T. Webb, Ph.D.

Virginia Commonwealth University
Richmond, Virginia
October 5th, 2012

Dedication

For my mother, whose passion for education and human rights will continue to inspire.

Carol Jean (Field) Peterson

February 2, 1948 - October 31, 2007

Acknowledgements

First and foremost, I wish to thank my advisor, Hermine H. Maes, Ph.D., for her patience, support and generosity with time and knowledge. It has truly been a pleasure to grow as a researcher under her guidance. I extend my sincere gratitude to my dissertation committee, Lindon J. Eaves, Ph.D., D.Sc., Kenneth S. Kendler, M.D., Brion S. Maher, Ph.D., Suzanne E. Mazzeo, Ph.D., Michael C. Neale, Ph.D., and Bradley T. Webb, Ph.D., for their mentorship and guidance throughout the dissertation process. In particular, I would like to thank Dr. Webb for his generosity with time and for graciously teaching me genetic association methods. I would also like to thank Dr. Neale for his generosity with grant funds, which allowed me to receive specialized training in statistical genetics, as well as the opportunity to present my research at numerous national and international conferences. I truly appreciate the exceptional education, training, and friendship I have gained from the faculty and students of the Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, and Departments of Mathematics and Biostatistics, including Drs. Tim Bigdeli, Jackie Meyers, Jia Yan, Rita Shiang, Brien Riley, Sam Chen, Shawn Latendresse, Tim York, Elizabeth Prom-Wormley, Nihar Sheth, David Bauer, RK Elswick and Leroy Thacker. I would also like to thank my first academic mentor, Nancy C. Raymond, M.D. of the University of Minnesota, an exceptional teacher, mentor, and friend, and a great inspiration. I would especially like to thank my family and friends. Without their love and support, I could not have pursued an advanced degree. To my parents, Tony and Carol Peterson, a very special thanks—for their unconditional love, encouragement and giving me every opportunity they could. They have given me the strength and foundation to strive to achieve my goals. I'd also like to thank my oldest friend and sister, Heidi Elmquist, for always being there for me and being a source of trusted, honest advice. Special thanks to Drs. Cynthia Barsness and Mary Ellen Bland for inspiring me from a young age to pursue the sciences and supporting my family through some of our most difficult times. Finally, a particular thanks to Dr. Tim Bigdeli, my best friend, for helping to keep me (partially) "sane" over the past five years and for the many hours of discussion and proof-reading that made this dissertation possible. It takes a village indeed—Vielen Dank!

Table of Contents

List of Tables

List of Figures

Global Abstract

Obesity is a serious public health crisis and recent estimates of its incidence are the highest in United States history, with 35% and 17% of American adults and children affected, respectively. The clinical definition of adult obesity is operationalized as a body mass index (BMI) greater than 30 kg/m^2. Although the prevalence of common obesity has increased dramatically over the past 30 years–largely thought to be due to changes in the environment, such as high calorie diets and sedentary lifestyles—twin and family studies have shown consistently that relative body weight is under considerable genetic influence in both children and adults, with heritability estimates ranging from 40% to 90%. Elucidating the genetic and environmental liability to relative body weight is an important public health endeavor. To further our understanding of the genetics of BMI and common complex obesity, several studies are described that integrate clinical, twin, and genome-wide association (GWAS) methodology in the context of genetic risk scores, clinical risk prediction, development across adolescence into adulthood, and comorbidity with depression symptoms and smoking behavior. First, in two cross-sectional genetic association studies, the utility of genetic risk sum scores (GRSS) were assessed, which summarize the total number of risk alleles, as an alternative form of replication and for potential clinical utility for obesity risk prediction. Next, since there has been only limited research on when during development BMI-associated variants begin influencing BMI, a longitudinal twin study was utilized to assess the effects of adult-validated BMI-SNPs across adolescence into adulthood. In addition, obesity is comorbid with numerous medical conditions including cardiovascular disease, insulin-resistance and some forms of cancer, as well as, various psychiatric disorders including eating disorders, mood disorders, and substance use. The next series of studies aimed to understand phenotypic and genetic associations between BMI/obesity and binge eating disorder (BED), depression symptoms and smoking behavior. Using a clinical sample of overweight and obese women with and without BED, the relationship of BED, food intake and internalizing symptoms of depression and anxiety was examined. Next, twin study methodology was used to investigate if shared genetic and/or environmental liability was responsible for phenotypic associations found between BMI, depression symptoms, and impulsivity. Finally, a genetic association study aimed at investigating whether genetic variants were associated with multiple behaviors, body composition and smoking behavior, or were trait-specific is presented. By utilizing several samples and methodologies and by pursuing methods development, a comprehensive approach is presented that is hoped to represent a more powerful evidence-based strategy to understanding the genetic and environmental determinants of BMI and common complex obesity, along with associated depression symptoms and smoking behavior.

Chapter 1: General Introduction

THE OBESITY EPIDEMIC

Prevalence

Obesity is a growing public health crisis that is increasingly global in scope (1). Its prevalence among adult Americans has increased dramatically over the last fifty years. As reported by the Centers for Disease Control and Prevention, obesity rates increased from 5% in 1959 to 15% in 1980, and recent estimates of 35% in 2010 are the highest in United States history (2). These estimates reflect a five-fold increase in obesity since 1959. Similarly, rates of obesity in childhood have increased significantly over the past 30 years, from 5% in 1980 to 17% in 2010 (2). Furthermore, the US is not alone in this epidemic, as the World Health Organization reports similar child and adult obesity trends for many other nations (1).

Defining obesity

Obesity is defined as an excess of body adiposity. Historically, body weight has been used as a proxy measure of adiposity. Until the 1970s, obesity was defined on the basis of reference tables of "ideal body weight" determined by the life insurance industry from associations with mortality (1). However, this was replaced in the 1980s by body mass index (BMI), a height-adjusted measure of weight calculated as the ratio of weight in kilograms by height in meters squared (kg/m^2). The current clinical definition of adult obesity is a BMI greater than 30 kg/m^2. BMI may be further partitioned into clinical categories corresponding to BMI ranges of underweight <18, normal 18-25, overweight 25-30 and obese class I 30-35, class II 35-40, and class III 40+ kg/m^2 (US Dietary Guidelines). In children, the criteria for classification as overweight and obese are based on the 85^{th} and 95^{th} percentiles of BMI for sex and age in relation to a reference population (3). Additionally, research has demonstrated that BMI is correlated with other, more direct measures of body fat including underwater weighing and dual energy X-ray absorptiometry (DEXA) (4-6). However, the limitations of BMI have been realized and factors such as age, sex, ethnicity, and muscle mass can affect the association between BMI and body fat (7-10). BMI nonetheless remains a widely used, simple, inexpensive, and noninvasive proxy measure of body fat that can be calculated with reasonable accuracy.

Mortality and morbidity

With increasing BMI there is a curvilinear rise in mortality (11, 12). In obese groups, this rise in mortality is thought to be due to the numerous adverse medical conditions associated with high levels of body fat. In adults, obesity is associated with increased risk of cardiovascular disease (13), type II diabetes (14, 15), some forms of cancer (16) and is

comorbid with multiple psychiatric disorders (17-21). Similarly, childhood obesity is associated with both immediate and long-term health consequences including increases in blood pressure, cholesterol and insulin resistance, as well as social and psychological problems (22-25). Furthermore, research has demonstrated that obese children are more likely to become obese adults (26-30), adding to the necessity of effective prevention and treatment efforts.

Determinants of obesity

Obesity is the result of positive energy balance, that is, excess caloric intake relative to energy expenditure. Although energy balance may appear straightforward, its relationship with obesity is complex and involves both genetic and environmental determinants. With respect to the doubling of obesity rates in the past 30 years, it is arguable that while our genomes have remained stable, environmental changes are attributable to this rise. Examples of such environmental factors include increases in restaurant and fast-food dining, and consumption of sweetened beverages (12, 31). For example, the reported number of fast food restaurants has increased from an estimated 600 in 1958 to over 222,000 in 2010 (1). Additionally, data on energy expenditure suggest that physical activity has declined but that the magnitude of this change is small and could not alone account for the dramatic increase in rates of obesity (32). Tracking of energy intake and expenditure is difficult and complicated by inaccuracies in reporting (33). Further research and developments in methodology are needed to clarify the relative contribution of dietary intake and energy expenditure to obesity over the life course. Although nutritional intake and physical activity affect relative body weight, twin and family studies have consistently shown a significant genetic contribution to body composition with heritability estimates of 40 to 70% (34-36). These results suggest that a considerable fraction of the variance in BMI is due to genetic effects. Therefore, the obesity epidemic likely reflects multiple interactions between lifestyle and genetic factors. More research is needed to unravel the interactions between these factors, and especially, identify critical time points of susceptibility.

THE GENETICS OF OBESITY

Three broad categories of obesity etiology have been described: monogenic, syndromic and common complex obesity. Dysfunction or loss of a single or few genes is both necessary and sufficient to cause monogenic or syndromic obesity and the typical onset is early in childhood. Common complex obesity is thought to be the result of the interplay between many genes, each of relatively small effect, along with influences of the environment. Research suggests that less than 5% of obesity cases are caused by monogenic or syndromic inheritance (37). Given that our focus herein is on common complex obesity, only a brief review of monogenic and syndromic obesity are provided. More detailed descriptions of rarer forms of obesity may be found in Mutch & Clement, Hinney *et al*., and Beales (38-40), as well as in the Online Mendelian Inheritance in Man database (http://www.ncbi.nlm.nih.gov/omim).

Monogenic obesity

Monogenic obesity, also known as non-syndromic obesity, is defined as obesity caused from a rare mutation of a single gene. There have been over 200 reported cases of monogenic obesity, implicating a total of 11 genes (38, 40). The most common mutations causing monogenic obesity are in the *melanocortin-4-receptor* gene (*MC4R*) on chromosome 18q22 and account for 6% of monogenic obesity (41, 42). Association with *MC4R* in humans was first reported in 1998 after screening of extremely obese individuals and their families identified frameshift mutations co-segregating in an autosomal dominant fashion (43, 44). The *leptin receptor* gene (*LEPR*) on chromosome 1p31 has been found to account for 3% of monogenic obesity cases (45) and yielded its first reported association in humans in 1998 following the presentation of severely obese siblings with extremely high levels of serum leptin (46). In fact, many of the transcripts of genes associated with monogenic obesity have been shown to have a role in the hypothalamic leptin-melanocortin system, which include the following genes: *leptin* (*LEP*), *pro-opiomelanocortin* (*POMC*), *prohormone convertase 1* (*PC1*), *brain-derived neurotrophic factor* (*BDNF*) and its receptor *neurotrophic tyrosine kinase receptor type 2* (*NTRK2*) and *single-minded homolog 1* (*SIM1*) (37, 38, 40, 47-50) .

Syndromic obesity

There are approximately 30 Mendelian disorders that include obesity as a clinical feature but are distinguished by additional presenting attributes including intellectual disabilities, dysmorphic features and developmental abnormalities (38, 40, 49, 50). These disorders are termed syndromic obesity and are the consequence of specific genetic defects or chromosomal abnormalities that disrupt contiguous gene(s). Because multiple genes may be disrupted, the particular causes of obesity often remain elusive. The most common forms of syndromic obesity disorders identified to date are Prader-Willi syndrome (PWS), Bardet-Biedl syndrome, and Alström syndrome. Of these, PWS has the greatest incidence, occurring in 1 in 25,000 births and, in addition to obesity, is characterized by hyperphagia, intellectual disabilities, and hypogonadism. Most cases of PWS are caused by deletion of the paternal copy of the imprinted *small nuclear ribonucleoprotein polypeptide N* gene (*SNRPN*) and potentially other genes within the 15q11-q13 region. The full catalogue of syndromes may be found in the Online Mendelian Inheritance in Man database (http://www.ncbi.nlm.nih.gov/omim).

Polygenic inheritance

Although the prevalence of common obesity has increased dramatically over the past 30 years–largely thought to be due to changes in the environment, such as high calorie diets and sedentary lifestyles—twin and family studies have shown consistently that relative body weight is under considerable genetic influence in both children and adults, with heritability estimates ranging from 40% to 90% (35, 51-54). Additionally, twin study meta-analyses which examined BMI from birth to adulthood have revealed that the contribution of genetic effects are low at birth but increase over time, with upwards of 50% of the phenotypic variance due to genetic effects after the first year of life (54, 55).

Furthermore, twin studies have demonstrated significant sex effects on BMI, with greater phenotypic variance in females and significant sex-specific genetic factors also reported (51, 54, 56-60). Given the large heritability estimates reported for BMI, molecular genetic approaches represent a useful tool with which to examine underlying mechanisms and genetic susceptibility to obesity. To date, a number of approaches have been utilized to identify BMI/obesity-associated genes including candidate gene, linkage and association studies.

Studies of candidate genes known to cause severe obesity in experimental animals have implicated several genes in human obesity (40). In the mid-1990s, Zhang *et al*., discovered that a mutation in the gene encoding the leptin protein was responsible for the severe obesity phenotype in the *ob/ob* mouse (61). Shortly thereafter, the first human mutations were reported in a pair of severely obese cousins, who were found to carry a frameshift mutation in the *LEP* gene on chromosome 7q32 (62). Mutations in the *LEP* gene are largely associated with monogenic obesity and ~1% of extreme early onset obesity cases carry *LEP* mutations (40). However, variants in *LEP* have not demonstrated association with BMI in the general population (63). Although hundreds of genes have been proposed as obesity candidate genes, few have yielded convincing association findings for BMI liability or obesity susceptibility and include common variants in *MC4R* and *BDNF* (63-67).

Genome-wide linkage studies provided an alternative method for identifying BMI/obesity-susceptibility genes. By examining rates of recombination between polymorphic markers among affected siblings, linkage analysis has the potential to localize a co-segregating genetic effect to a particular genetic locus. Unlike candidate gene studies, linkage studies do not rely on an *a priori* hypothesis, but rather aim to identify previously unknown genetic loci to potentially lead to new insights regarding the biology. Numerous linkage scans have been performed, identifying more than 300 chromosomal loci demonstrating linkage with BMI/obesity (48, 68). However, like candidate gene approaches, linkage studies have been plagued by non-replication of positive findings (68, 69). For instance, a meta-analysis of 37 studies boasting a combined sample size of 10,000 families failed to identify any locus robustly linked to BMI or obesity (68). As such, linkage analysis has not proved to be a powerful method for identifying genetic loci with small effects, as would be expected for BMI and common complex obesity.

By the mid-2000s, the fruits of the Human Genome Project and International HapMap Project (70), coupled with the rapid development of high-density high-throughput genotyping arrays, set the stage for a new era of complex disease mapping by genome-wide association studies (GWAS). GWAS is premised on the expectation that, by capturing the majority of common human variation across the genome, individual associations might be identified without *a priori* expectation of a given locus's involvement in disease etiology. Common variation, in the context of GWAS, is taken commonly to mean point mutations, or single nucleotide polymorphisms (SNPs) with minor allele frequencies (MAF) >1-5%. An advantage of GWAS over linkage studies is its extendibility to population-based designs, allowing for potentially larger sample sizes and increased power to detect variants with smaller effect-sizes (71). The GWAS approach has successfully identified polymorphisms that contribute to disease risk for numerous complex traits and diseases (72). Though, in some ways the field of obesity

stumbled into the GWAS era. In 2007, a GWAS of type II diabetes by the Welcome Trust Case Control Consortium identified the first association with BMI. A SNP in the *fat mass and obesity-associated* (*FTO*) gene was found to be significantly associated with type II diabetes. However, when the analyses were adjusted for BMI, the strength of this association was diminished, indicating that the effect of *FTO* on type II diabetes was through its association with BMI. Since 2007, several subsequent GWAS of successively larger size have been performed for BMI and obesity-related traits (67). In 2009, two large-scale BMI meta-analyses by Thorleifsson *et al.* and Willer *et al*. yielded 13 genetic loci reaching genome-wide significance, including the previously implicated variants in or near *FTO* and *MC4R* (see Chapter 2 for a complete list). A subsequent mega-analysis by Speliotes *et al.* (2010) incorporated a two-stage approach in which a GWAS was performed on 249,796 individuals from 46 studies in the first stage, followed by a second stage in which association was performed in an additional 125,931 individuals from 42 studies. This study confirmed 32 SNPs unequivocally associated with BMI (see Chapter 3 for a complete list). These variants, although highly associated with BMI, have small individual effects ranging 0.06 to 0.39 kg/m^2 change in BMI per risk allele and in aggregate account for a limited proportion of the phenotypic variance (~1.45%) (63).

Current GWAS designs are limited to detecting trait or disease associations with common variation in accordance with the common disease-common variant (CDCV) hypothesis (73). For BMI, the aforementioned 32 common SNPs account for ~1.45% of the phenotypic variance, leaving a substantial fraction of the heritability in BMI unaccounted for. As for other complex traits and diseases, this "missing heritability" has lead to efforts to identify rare variants contributing to common disease. Given the heritability of BMI and the observation that common SNPs only account for a portion of the expected phenotypic variance, it is conceivable that additional classes of genetic variants such as rarer and/or structural variation or epigenetic mechanisms influence body composition. A growing number of rare copy number variants (CNV) have demonstrated association with BMI and obesity (a catalogue of CNVs appears in Chapter 3) (74-82). In addition, for many of the BMI/obesity-associated loci, it has yet to be determined if they represent the causative locus or if they are merely correlated with the causative variant. Fine mapping efforts by large-scale exome and genome sequencing efforts are needed to identify the true causal variants. Indeed, such studies are underway and include the UK10K project, a whole-genome sequencing study of 4,000 individuals and exome sequencing of an additional 6,000 individuals, including 2,000 with extreme obesity phenotypes (83).

In summary, despite an arguably changing environment, twin and family studies support the significant role of genes in contributing to relative body weight and obesity across the lifespan. However, as described in the preceding sections, most genes that contribute to relative body weight and obesity are of largely unknown function and have limited utility for risk prediction. This is further complicated by the fact that most studies to date have been on samples of primarily European descent and cross-sectional in nature. Additional research is needed in diverse human populations and it remains unknown when in development the identified genetic effects become important for predicting BMI.

OBESITY AND PSYCHIATRIC COMORBIDITY

Obesity is comorbid with numerous medical conditions including a variety of psychiatric disorders and traits. Obesity has been associated with eating disorders, mood disorders, substance use as well as personality disorders (84-86). For example, the National Epidemiological Survey on Alcohol and Related Conditions (NESARC), a study of over 40,000 American adults, reports significant increased odds for many psychiatric disorders among them include: lifetime prevalence of any anxiety disorder (OR = 1.4-2.3), lifetime prevalence of alcohol dependence (OR = 1.1-1.6), and prevalence of antisocial personality disorder (OR = 1.1-3.3) (86). Below appears a synopsis of three psychiatric disorders and their association with obesity that are of particular relevance to this thesis.

Binge eating disorder

Binge eating disorder (BED) is under consideration for inclusion in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V). BED is currently defined by the DSM-IV as a provisional eating disorder diagnosis characterized by recurrent episodes of binge eating without weight control compensatory behavior and includes: (1) "eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than what most people would eat during a similar period of time and under similar circumstances," and (2) "a sense of lack of control over eating during the episode". Additionally, individuals with BED must experience distress about their binge eating and endorse three of the following symptoms: (1) eating more rapidly than normal, (2) eating until uncomfortably full, (3) eating large amounts when not hungry, (4) eating alone because of embarrassment, and (5) feeling disgusted, depressed or guilty about overeating (87). Although obesity is not a requirement for a BED diagnosis, research indicates that approximately 70% of those meeting criteria for BED are obese (21). While the prevalence of BED in community samples ranges from 2-5%, approximately 30% of obese individuals seeking weight control treatment meet criteria for BED (88, 89). The recurrent overeating that characterizes BED, along with the absence of compensatory behaviors exhibited by those with bulimia nervosa (BN), is most likely responsible for the high frequency of obesity in this group.

Major depressive disorder and depression symptoms

According to the DSM-IV, major depressive disorder (MDD) is characterized by a depressed mood most of the day nearly everyday for at least a two-week period and/or diminished interest or pleasure in all or almost all activities. Additional criteria include endorsement of at least three of the following symptoms: (1) significant weight loss or weight gain, (2) insomnia or hypersomnia, (3) psychomotor agitation or retardation, (4) loss of energy, (5) feelings of worthlessness or excessive guilt, (6) diminished ability to concentrate and (7) recurrent thoughts of death or suicide (87). As reported by the 2006 National Comorbidity Survey Replication, the lifetime history estimates of MDD are 12.7% in men and 21.3% in women (90). However, within obese populations, reported lifetime prevalence rates of depression have been shown to be elevated upwards of 32% (20). In addition, Strine *et al.* found that adults with a current or lifetime diagnosis of

depression were significantly more likely to engage in unhealthy behaviors such as physical inactivity and to be obese (20). Furthermore, longitudinal phenotypic studies have found a reciprocal association between obesity and depression, suggesting that elevated BMI may increase depression and vice versa (91, 92). Cross-sectional studies of BMI and depression symptoms have reported positive (93-97), negative (primarily in males) (98, 99) and no association (100-102) between these traits. However, a population based study from the Netherlands found a quadratic (U-shaped) association of BMI and depression indicating those with the lowest and the highest relative body weight were more likely to present with depression. In light of current DSM-IV MDD criteria, which include items related to increase and decrease in appetite, weight and energy expenditure, it is feasible that BMI may be associated with greater levels of depression in both underweight and obese individuals (103). Further research is needed to clarify the nature of the association between body weight and depression.

Nicotine dependence and smoking behavior

Nicotine dependence (ND) is characterized by tolerance and withdrawal symptoms in relation to tobacco use. ND can occur with cigarette smoking, smokeless tobacco use, cigar or pipe use. According to the DSM-IV, ND is diagnosed by clinically significant impairment or distress from the presence of any three of the following seven criteria occurring at any time in the same 12-month period: (1) tolerance, (2) withdrawal, (3) taking the substance in larger amounts than intended, (4) persistent desire or unsuccessful efforts to cut down on the substance, (5) spending a great deal of time obtaining or recovering from the effects of the substance, (6) giving up important recreational, social, or occupational activities as a result of the substance, and (7) continued use of the substance despite physical or psychological problems caused by the substance. There are a number of questionnaires that are used to assess ND and the most widely used are the eight-item Fagerström Tolerance Questionnaire (FTQ), the six-item Fagerstöm Test for Nicotine Dependence (FTND), which is a shortened version of the FTQ, excluding items on nicotine yield of cigarettes and inhalation, and the two-item Heaviness of Smoking Index (HSI), a shorter version of the same test only including items on time to first cigarette after waking and number of cigarettes per day (104, 105). In 2010, according to the Centers for Disease Control and Prevention, the estimate of American adults reported as current smokers was 19.3% (106-108). Cross-sectional studies of smoking behavior typically support a negative relationship between current smoking and BMI (109-111), which may be due, in part, to effects of nicotine on energy homeostasis (112-116). Furthermore, smoking cessation is often followed by weight-gain (113, 117, 118). In contrast, however, a positive association is supported by the observation that within smoking cohorts, heavy smokers tend to be of increased body weight compared to light smokers (119-121). Additionally, smoking has been associated with accumulation of visceral fat and increased waist circumference (122-124). Phenotypic associations between smoking and body composition suggest a complex relationship and the causes of these associations remain incompletely understood.

SPECIFIC AIMS

Survey of limitations

GWAS has proven a fruitful approach for identifying polymorphisms that contribute to disease risk for numerous complex traits and diseases (72). However, this method has been met with important limitations, especially as applied to psychiatric disorders. A number of potential factors have been proposed that may reduce the power of this methodology in general, as well as for the field of common complex obesity specifically.

To date, large-scale GWAS meta-analyses have confirmed 32 SNPs associated with BMI which, although highly associated, have small individual effects ranging 0.06 to 0.39 kg/m^2 change in BMI per risk allele. Therefore, replication attempts have limited power to achieve genome-wide significance, even with thousands of subjects (125). Moreover, in aggregate these BMI-associated SNPs account for a fraction of the phenotypic variance (~1.45%) (63), and thus have limited utility for risk prediction (67), suggesting that other classes of genetic variants may be important.

Also, given that the large-scale meta-analyses of BMI were performed on samples of primarily European descent, these findings may not be easily generalizable to other ancestry groups. From a public health perspective, this is particularly problematic since research indicates that there exist health disparities between racial groups, including increased obesity prevalence in African- and Hispanic-Americans (126). The aforementioned BMI-associated SNPs were identified from cross-sectional adults samples, which does not address at what point during development these variants influence BMI. The identification of specific "windows" of risk is essential for understanding development as well as informing prevention and intervention efforts.

Furthermore, relative body weight has been associated with numerous other medical conditions and traits. This may impact the power of gene identification efforts, especially if control groups are not adequately screened for correlated traits or such correlations are not accounted for in statistical methodology. There is a paucity of literature reporting on the potential common genetic liability between obesity and comorbid traits. Without consideration of genetically correlated traits, genome-wide studies of complex disease may be limited in their power to detect etiologically relevant variation.

In summary, the present survey of limitations of gene-identification efforts for common complex obesity has identified the following issues: replication of variants with small effects, utility of risk prediction, generalizability to multiple racial groups and across the lifespan and affects of comorbidity. This thesis delves into many of these limitations and attempts to address these issues through five specific aims.

Specific aims

The purpose of this research was to develop methods to better delineate the genetics of common complex obesity and the corresponding associations with depression symptoms and smoking behavior through the following five aims:

1. Examine phenotypic associations between BMI, depression and smoking behavior in clinical and epidemiological samples. Identify putative mediators and moderators of the BMI-depression link and explore sample structures via symptom profiles.

2. Apply multivariate twin methodology to BMI, depression and nicotine-use phenotypes in order to test for shared genetic and environmental liability of multiple traits and stability over time.

3. Catalogue common polygenic variation associated with body composition.

4. Test genetic variants catalogued for association in multiple cohorts and traits to provide evidence of replication, assess clinical utility and potentially discover variants influencing multiple traits.

5. Methods development in each of the preceding areas, presented throughout.

Thesis outline

In the following chapters, several studies will be described that integrate both GWAS and twin study methodology to further our understanding of the genetics of BMI and common complex obesity in the context of genetic risk scores, clinical risk prediction, development across adolescence into adulthood, and comorbidity with depression symptoms and smoking behavior. In the first chapter, the obesity epidemic and the associated mortality and morbidity, and a highlight of the genetics of obesity and BMI are reported in order to provide the necessary background.

In the subsequent two chapters, genetic risk sum scores (GRSS), which summarize the total number of risk alleles and test the aggregate risk, as an alternative form of replication and assess clinical utility for obesity risk prediction are performed. Specifically, in Chapter 2, genetic variants were catalogued from two-large scale meta-analyses of BMI in order to test a GRSS constructed by the count method in a sample of European-Americans and African-Americans from the Molecular Genetics of Schizophrenia Controls (MGS-C). In Chapter 3, to extend GRSS methodology, scores were constructed from proxy versus imputed SNPs and count versus weighted methods were compared. In addition to BMI-validated SNPs, previously implicated common and rare CNVs were identified from the literature and were tested for association with BMI and obesity. An integrated model of common and rare variation was tested for association with BMI and subsequently assessed for clinical utility in a sample of European-Americans and African-Americans from the Study of Addiction: Genes and Environment (SAGE).

Since there has been limited research on when during development BMI-associated variants begin to influence BMI, Chapter 4, utilizes a longitudinal twin study to assess the effects of adult-validated BMI-SNPs across adolescence into adulthood. To our knowledge, this is the first study of BMI to incorporate GRSS methodology in the context of variance decomposition. Furthermore, only limited models have been applied to examine the genetic and environmental architecture of BMI across adolescence into adulthood. Therefore, this study tested models to quantify the relative proportion of genetic and environmental factors that persist across time versus those that are time specific in the Virginia Twin Study of Adolescent Behavioral Development (ABD).

Obesity is comorbid with numerous medical conditions as well as a various psychiatric disorders including eating disorders, mood disorders, and substance use (84-86). In Chapters 5 through 8, phenotypic and genetic associations between BMI/obesity and binge eating disorder (BED), depression symptoms and smoking behavior are examined in several different types of samples. In Chapters 5 and 6, the University of Minnesota Study of Binge Eating Disorder (UofMN), which is a clinical sample of overweight and obese women with and without BED, is used to examine the relationship of BED, food intake and internalizing symptoms of depression and anxiety. Additionally, tracking of energy intake and expenditure is difficult and complicated by inaccuracies in reporting (33). An improved understanding of the accuracy of self-reported food intake is central to diagnosis of eating disorders, monitoring response to treatment and obesity management. Therefore, in Chapter 5, energy intake and energy expenditure were assessed by multiple methods to potential identify differences in food intake, metabolism and accuracy of self-reported food intake in obese women with and without BED. In Chapter 6, the UofMN sample was used to examine models by which BED, internalizing behaviors of depression and anxiety influence food intake in overweight/obese women. Greater understanding of the mechanisms underlying the associations between mood, binge eating and food intake will facilitate the development of more effective prevention and treatment strategies for both BED and obesity.

Despite numerous phenotypic associations between BMI, depression symptoms, and smoking behavior, there is a paucity of reports investigating genetic and environmental associations between them. To better understand the underlying common genetic architecture, it is essential that the complex nature of the observed associations between these traits be assessed. Accordingly, Chapters 7 and 8 investigate associations between BMI, depression symptoms and smoking behavior by two different types of genetically informed samples: twin studies and GWAS. In Chapter 7, twin study methodology is used to investigate if shared genetic and/or environmental liability is responsible for phenotypic associations found between relative body weight, depression symptoms, and impulsivity in the Virginia 30,000 Twin Study (VA30k). Furthermore, most studies do not examine common versus specific genetic effects. In Chapter 8, genetic variants individually associated with BMI or smoking behavior were catalogued and tested for association in The Health Aging and Body Composition Study (HABC) in order to investigate whether genetic variants were associated with multiple behaviors or were trait-specific. Without consideration of genetically correlated traits, genome-wide studies of complex disease may be limited in their power to detect etiologically relevant variation.

Finally, Chapter 9 provides a global discussion of this thesis by summarizing key findings from each study, discussing limitations of the research presented, and providing extensions for future research. By utilizing several samples and methodologies and by pursuing methods development, a comprehensive approach is presented that is hoped to represent a more powerful evidence-based strategy to understanding the genetic and environmental determinants of BMI and common complex obesity, along with the associated depression symptoms and smoking behavior.

COHORTS

In order to implement the aims of this dissertation, several data types were utilized including clinical studies, population-based twin samples and samples genotyped for the study of common diseases. Table 1 lists each study along with the corresponding sample and the chapter in which it appears.

I. Molecular Genetics of Schizophrenia Controls (MGS-C)

The MGS-C is a cross-sectional sample of 2,653 European-Americans and 973 African-Americans. Participants were considered for "control" status if they denied all of the following psychosis screening questions: treatment for or diagnosis of schizophrenia or schizoaffective disorder; treatment for or diagnosis of bipolar disorder or manic-depression; treatment for or diagnosis of psychotic symptoms such as auditory hallucinations or persecutory delusions. Participants completed an online questionnaire, which included items on height and current weight. Venipuncture for DNA extraction and establishment of lymphoblastoid cell lines was completed at Rutgers University Cell and DNA Repository. DNA samples were genotyped using the Affymetrix 6.0 array at the Broad Institute.

II. Study of Addiction: Genes and Environment (SAGE)

Complete data on height, weight, alcohol dependence, nicotine dependence, genotypes and copy number variants were available for 1850 European-American and 498 African-American SAGE participants. The SAGE sample was drawn from three contributing projects: the Collaborative Study on the Genetics of Alcoholism (COGA), the Collaborative Study on the Genetics of Nicotine Dependence (COGEND) and the Family Study of Cocaine Dependence (FSCD). Body composition variables were not available for the FSCD sample, thus not included in the analyses described herein. Study variables were assessed by interview using versions of the Semi-Structured Assessment for the Genetics of Alcoholism. Body mass index (BMI) was calculated from self-reported height and weight. Samples were genotyped on the Illumina Human 1M beadchip at the Center for Inherited Diseases Research at Johns Hopkins University.

III. *Virginia Twin Study of Adolescent Behavioral Development (ABD)*

ABD is a longitudinal population-based twin study of adolescent psychopathology. ABD twin participants twins aged 8 to 17 were recruited through Virginia schools and were followed-up every 18 months for up to four waves of data collection (n = 2,794). In total, there were 913 participants from 639 families (291 twin pairs, 348 singletons) that were also genotyped on the Illumina Human 660 array. BMI was calculated from measured weight and height.

IV. *University of Minnesota Study of Binge Eating Disorder (UofMN)*

Thirty-four women participated in this study examining metabolic measures, including 17 meeting clinical criteria for binge eating disorder and 17 overweight/obese controls with no history of any binge eating or eating disorder behaviors. Food intake was assessed from a laboratory eating episode, 24-hour dietary recall interviews and food diaries. Energy expenditure was assessed by the doubly labeled water technique, basal metabolic rate and the thermic effect of food. Furthermore, participants completed a variety of questionnaires including the Beck depression and anxiety inventories.

V. *The Virginia 30,000 Twin Study (VA30k*)

The VA30k sample is a large population-based twin study. Ascertainment for the VA30k sample was through two sources, a volunteer twin sample solicited through the American Association of Retired Persons and the Virginia Twin Registry. BMI data was available for n=14,457 adult twins. Participants completed the Health and Lifestyle Questionnaire, which included abbreviated versions of the Symptoms Checklist and the Eysenck Personality Questionnaire and smoking history.

VI. *The Health Aging and Body Composition Study (HABC)*

The HABC study is a prospective community based sample of body composition changes over time in elderly adults and included 1663 European-American and 1139 African-American participants. Participants were recruited in 1997-1998 from Pittsburgh, PA, and Memphis, TN metropolitan area residents who were Medicare eligible and between the ages of 69 and 80 years. BMI was calculated from laboratory measured height and weight during initial evaluation. Physical activity was estimated from a structured interview of 27 questions. Computerized tomography was used to determine abdominal visceral adiposity density. Smoking habits and race were self-reported via telephone interview. Genotyping was performed by the Center for Inherited Disease Research using the Illumina Human 1M-Duo BeadChip system.

Table 1: Summary of dissertation studies

Chapter	Study	Sample	Design	Phenotype	Aim
2	Genetic risk sum score comprised of common polygenic variation is associated with body mass index	MGS-C	Cross-sectional, GWAS	BMI, obesity	3, 4, 5
3	Association of common and rare variation influencing body mass index: A combined single nucleotide polymorphism and copy number variation analysis	SAGE	Cross-sectional, GWAS, CNV	BMI, obesity	3, 4, 5
4	Association of common polygenic variation with body mass index across adolescent development: A longitudinal twin study	ABD	Longitudinal, twin, GWAS	BMI	2, 4, 5
5	Comparisons of energy intake and energy expenditure in overweight and obese women with and without binge eating disorder	U of MN	Clinical	BED, obesity	1, 5
6	Binge eating disorder mediates links between symptoms of depression, anxiety, and energy intake in overweight and obese women	U of MN	Clinical	BED, obesity, depression symptoms	1, 5
7	Genetic and environmental associations between body mass index, depression symptoms and impulsivity in a population-based sample of twins: VA30k	VA30k	Cross-sectional, twin	BMI, depression symptoms, impulsivity	1, 2
8	On the genetic and environmental relationship of body mass index, smoking initiation and nicotine dependence in a population-based sample of twins	VA30k	Cross-sectional, twin	BMI, smoking behavior	1, 2
8	Evidence of shared polygenic risk among smoking behaviors and body composition	HABC	Cross-sectional, GWAS	BMI, obesity, smoking behavior	1, 3, 4

Chapter 2: Genetic risk sum score comprised of common polygenic variation is associated with body mass index

Adapted from: Peterson RE, Maes HH, Holmans P, Sanders AR, Levinson DF, Shi J, Kendler KS, Gejman PV, Webb BT. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Human Genetics*. 2011 Feb;129(2):221-30.

ABSTRACT

Genome-wide association studies (GWAS) of body mass index (BMI) using large samples have yielded approximately a dozen robustly associated variants and implicated additional loci. Individually these variants have small effects and in aggregate explain a small proportion of the variance. As a result, replication attempts have limited power to achieve genome-wide significance, even with several thousand subjects. Since there is strong prior evidence for genetic influence on BMI for specific variants, alternative approaches to replication can be applied. Instead of testing individual loci sequentially, a genetic risk sum score (GRSS) summarizing the total number of risk alleles can be tested. In the current study, GRSS comprised of 56 top variants catalogued from two large meta-analyses was tested for association with BMI in the Molecular Genetics of Schizophrenia controls (2,653 European-Americans, 973 African-Americans). After accounting for covariates known to influence BMI (ancestry, sex, age), GRSS was highly associated with BMI (p-value = 3.19×10^{-6}) although explained a limited amount of the variance (0.66%). However, area under receiver operator criteria curve (AUC) estimates indicated that the GRSS and covariates significantly predicted overweight and obesity classification with maximum discriminative ability for predicting class III obesity (AUC=0.697). The relative contributions of the individual loci to GRSS were examined *post hoc* and the results were not due to a few highly significant variants, but rather the result of numerous variants of small effect. This study provides evidence of the utility of a GRSS as an alternative approach to replication of common polygenic variation in complex traits.

INTRODUCTION

Obesity is a general medical condition, defined clinically by a body mass index (BMI) greater than 30 kg/m^2 and is associated with increased risk of cardiovascular disease, type II diabetes, cancer and poor quality of life (12, 127, 128). The National Center for Health Statistics reports over 33% of American adults are obese with another 33% meeting criteria for being overweight (127, 129). Although increase in energy intake with reduced physical activity contributes to the increase in obesity, genetic factors have consistently been demonstrated to influence individual differences in BMI, with twin and family studies estimating heritabilities of ~0.70 (35, 36).

Genome-wide association studies (GWAS) have successfully identified polymorphisms that contribute to disease risk for numerous complex traits and diseases (72). GWAS for BMI and obesity using sample sizes in the tens of thousands have yielded many putative risk variants of individually small effect. The first common single nucleotide polymorphisms (SNPs) associated with BMI and common obesity were in the fat mass and obesity-associated (*FTO*) gene and near melanocortin 4 receptor (*MC4R*) and have been widely replicated (66, 130-135). Additionally, two large-scale BMI meta-analyses, Thorleifsson *et al.* (2009) and Willer *et al*. (2009), yielded 13 genetic loci reaching genome-wide significance, including the previously implicated variants in or near *FTO* and *MC4R*. These variants were highly significant but had modest effects with 0.06-0.4 kg/m^2 per allele change in BMI and modest obesity (BMI>30 kg/m^2) odds ratios ranging 1.03-1.3. Although many loci are expected to contribute to a complex trait like BMI, the large number implied by the current result was unexpected to many (136, 137). Despite the large sample size (n>30,000), Willer *et al.* (2009) estimated 5-10% power to detect genome-wide significant variants with effect sizes of 0.06-0.1 BMI units per allele. Therefore, it is likely that many variants influencing BMI did not reach genome-wide significance in these meta-analyses.

Replication attempts using studies unselected for BMI have limited power to achieve genome-wide significance, even with thousands of subjects (125). Since there is strong *a priori* evidence for genome-wide significant and suggestive variants from the large meta-analyses, alternative approaches to replication can be applied. Instead of testing individual loci sequentially, a genetic risk sum score (GRSS) summarizing the total number of risk alleles can be constructed and tested. The aggregate risk should be significant if a sufficient proportion of the variants have real effects. GRSS have been used to test the total impact of associated variants on complex traits and disease. For example, Aulchenko *et al*. (2009) used 54 variants in a GRSS which accounted for ~4% of the phenotypic variance in height. Risk scores incorporating 18-20 genome-wide significant variants have been shown to be associated and predictive of type II diabetes, though algorithms including family history and additional risk factors perform better (138, 139). GRSS have also been applied to BMI and obesity in populations of European and Chinese descent which incorporated 8-15 variants and accounted for 0.5-1.12% of the phenotypic variance (64, 65, 140-143). Presently, BMI GRSS have only incorporated genome-wide significant variants. However, research by Evans *et al*.(2009), suggests that in some cases, including bipolar disorder, coronary heart disease, hypertension and type

II diabetes, using liberal thresholds ($\alpha = 0.5$) for SNP selection in GRSS may improve predictive ability.

The purpose of this study was to test a GRSS comprised of replicated genome-wide significant variants as well as additional variants with suggestive evidence catalogued from large scale meta-analyses for association with BMI in 2,653 European-Americans and 973 African-Americans from the Molecular Genetics of Schizophrenia control sample (MGS-C). Based on the expected BMI effect sizes of 0.05-0.3 kg/m^2 per allele change in BMI, the MGS-C sample would have limited power to detect genome-wide significant variants for individual loci. However, the aggregate risk should be adequate if a sufficient proportion of the reported variants are real. Therefore, these analyses serve as a replication attempt of top variants catalogued from large-scale meta-analyses via a sum score approach.

MATERIALS AND METHODS

Participants and phenotypes

The MGS-C sample has been previously described in detail (144-146). In summary, Knowledge Networks, Inc., a survey research company, recruited self-identified non-Hispanic European-American and African-Americans from a nationwide panel of survey participants, which was assembled by random digit dialing except 772 of the African-Americans were recruited through a subcontract to Survey Sampling International by internet banner ad recruitment. The institutional review board approval was obtained at NorthShore University HealthSystem and participants completed an online consent with an identical hard-copy consent signed at venipuncture. Participants completed an online questionnaire, available at nimhgenetics.org, which included items on height and current weight. BMI was calculated from respondents' self reported height and current weight. Participants were removed from data analysis if there was missing data on either height or weight or if calculated BMI was less than 15 or greater than 60 as values not in this range were likely data entry errors. There were 2,653 European-Americans and 973 African-Americans included in the present study. Phenotypic details are displayed in Table 2 with full sample characteristics found in Sanders *et al.* (146).

Genotyping

Venipuncture for DNA extraction and establishment of lymphoblastoid cell lines was completed at Rutgers University Cell and DNA Repository. DNA samples were genotyped using the Affymetrix 6.0 array at the Broad Institute. There were 3,827 participants genotyped (n=2,817 European-American, n=1,010 African-American) of which 3,626 (95%) passed stringent quality control criteria. Principal component (PC) scores reflecting continental and within-Europe ancestries of each subject were computed and outliers were excluded. Genomic control λ values for autosomes after quality control procedures were 1.005 for African-American and 0.998 for the European-Americans.

Selection of 56 SNPs

Preliminary SNP selection identified 78 variants meeting criteria for genome-wide or suggestive significance in either of two large meta-analyses of BMI, 43 from Thorleifsson *et al.* (2009) and 35 from Willer *et al.* (64, 65). Thorleifsson and colleagues report genome-wide significant ($p < 1.6x10^{-7}$) associations with 29 SNPs in 11 chromosomal regions, using a discovery sample of n=34,416 and replication sample of n=5,586. The Willer *et al.* meta-analysis detected 8 genome-wide significant ($p<5.0x10^{-8}$) SNPs in first- and second-stage samples of n=32,387 and n=54,316, respectively. Only variants in or near FTO and MC4R were found to be genome-wide significant in both meta-analyses. The remaining genetic loci were suggestive in the opposing meta-analyses ($p<0.05$) except rs7138803 on 12q13 ($p=0.14$). Significance level for one SNP, rs10938397 on 4p12, could not be compared between meta-analyses because there was no corresponding proxy SNP. Of the 78 variants catalogued, 29 had matching SNPs on the Affymetrix 6.0 array. For the 49 SNPs not present, proxies (45 r2>0.8; 4 r2>0.7) were identified using SNP Annotation and Proxy Search V2.1 (147). Following removal of 7 duplicate proxies and 6 variants from Willer *et al.* for which no proxies were available (r2>0.7), 65 SNPs remained. Haploview version 4.10 was used to determine phase and corresponding proxy alleles (148, 149). In order to avoid bias due to correlated effects, SNP pruning ($r^2>0.8$) was performed using PLINK v. 1.07p (150). Of the 56 remaining SNPs, 19 met genome-wide significance criteria in the two meta-analyses. The additional 37 were included as they were the next top SNPs reported ($p<0.05$). Although our SNP selection threshold was more liberal than the traditional genome-wide significance threshold, it was more conservative than other models of complex disease risk prediction (151, 152). Table 5 details information on the 78 catalogued SNPs.

Genetic risk sum score

Under an additive model, 56 variants were used to construct the GRSS. The use of an additive model was chosen as specific non-additive effects have yet to be associated and confirmed in the literature. The GRSS was calculated by summation of the number of risk alleles across the 56 variants divided by the number of SNPs in the score to obtain an average number of risk alleles per locus. GRSS were calculated using the profile option in PLINK. If SNP information was missing in an individual then the scoring routine imputed expected values based on sample allele frequency. R version 2.20.0 was used to fit linear regression models using standard covariates and GRSS as predictors with BMI as the outcome variable. To facilitate interpretation of effects in linear models independent variables were centered.

Prediction of obesity

One method to assess diagnostic efficiency is to graph a receiver operator criteria (ROC) curve, which is a plot of the true positive rate (sensitivity) against the false positive rate (1 - specificity) and calculate the corresponding area under the curve (AUC). An AUC may range from 0.5, non-informative, to a maximum of 1.0, perfect discrimination between cases and controls. An AUC is the probability that the predictor is greater for

cases than controls (153, 154). Generally, an AUC of 0.80 is suitable for screening while 0.99 is acceptable for diagnosis (155). To test various BMI thresholds, current BMI was dichotomized to create categories of overweight and obesity class I, II and III which had corresponding ranges of BMI> 25, 30, 35 and 40 kg/m^2 respectively. Binary logistic regression was used to calculate predicted probabilities of the models and was used as the predictor to generate ROC curves. Discriminative accuracy of the GRSS and covariates (molecularly derived ancestry, sex, age, ancestry by sex interactions) to predict BMI category was estimated by calculating the AUC from ROC curves using PASW Statistics version 17.0.

RESULTS

Phenotypic detail

Descriptive statistics for age and BMI are presented by race and sex in Table 2. The mean age of participants was 48.8 and ranged from 18 to 90 and as depicted in Figure 3 produced a relatively normal distribution. BMI was not significantly associated with age ($p=0.135$, Figure 4). Males were significantly older than females and European-American females and males were significantly older than African-American females and males ($p<0.0001$). When partitioning the sample by clinically established BMI (kg/m^2) categories, 29.0% was either under or normal weight (BMI<25), 33.4% was overweight (25≤BMI <30), 20.4% was obese class 1 (30≤BMI<35), 9.5% was obese class II (35≤BMI<40) and 7.7% was obese class III (40≤BMI). There was a significant ancestry by sex interaction with BMI. As expected, females had significantly greater BMI than males with African-American females having greater BMI than European-American females and African-American males having greater BMI than European-American males ($p<0.0001$). Phenotypic findings in the MGS-C sample are consistent with cross-sectional data from the National Center for Health Statistics and National Health and Nutrition Examination Study (156), finding obesity more prevalent in women and African-Americans. Additional sample characteristics have been previously reported (146).

Genetic risk sum score

Fifty-six variants catalogued from two large-scale BMI meta-analyses were used to construct the GRSS (64, 65). These variants were summarized in the GRSS, which was calculated by the summation of the number of risk alleles across the SNPs for each individual divided by the number of SNPs in the score to achieve an average allele count. The frequencies of GRSSs are shown in Figure 1 and produced a relatively normal distribution. The mean GRSS, or average number of risk alleles present per locus, was 0.494 (SD=0.052) with a range from 0.318 to 0.691, which corresponds to an average of 55 risk alleles per person with a range from 35 to 77.

Results from linear regression analyses are presented in Table 3. Standard covariates known to influence BMI (ancestry, sex and age) were included in the models. Described previously (144, 145), 224 ancestry informative markers were used to construct 10 PC scores designed to discriminate between European, African, Ameri-

Indian and Asian ancestry. PC1 (distinguishes European from African ancestry) and PC4 (distinguishes Eastern and Western European ancestry) were significantly associated with BMI and therefore included as covariates. Interactions between the covariates were tested and significant interactions were found between PC1 and sex and PC4 and sex. No other interactions between the covariates were significant. Model 1, the base model, included the standard covariates and the significant interactions between ancestry PCs and sex and accounted for 3.5% of the variance in BMI. Model 2, which added the GRSS to the base model, fit significantly better ($F(1,3027) = 21.8$, $p = 3.2 \times 10^{-6}$) and accounted for an additional 0.66% of phenotypic variance in BMI for a total of 4.1%. We note that the GRSS accounted for more of the variance in BMI than either sex or age. Interactions between the covariates and the GRSS were tested but no significant interactions were found (presented in Table 6). Therefore, our results suggest that GRSS was equally associated with BMI in men and women, in European- and African-Americans and across all ages. The relative contributions of the individual loci to the GRSS were examined post hoc by dropping the most significantly associated SNP from the score iteratively until the score was no longer statistically associated with BMI. As depicted in Figure 2, the GRSS reached non-significance after dropping the top 23 variants.

Prediction of obesity

To test the discriminative accuracy of the GRSS and covariates (molecularly derived ancestry, sex, age, ancestry by sex interactions) to predict obesity, ROC curves were plotted and the corresponding AUC were calculated. To test various BMI thresholds, current BMI was dichotomized to create categories of overweight and obesity class I, II and III. Figure 2 displays statistics from ROC curve analysis by BMI category. AUC estimates indicated that the model significantly predicted overweight and obesity classification with maximum discriminating ability when predicting class III obesity (AUC=0.697, 95% CI= [0.663, 0.731]). We note that the clinical setting may prefer to use self-identified ancestry as opposed to molecularly derived ancestry in risk prediction because of genotyping cost. In the MGS-C data, the use of self-identified ancestry did not greatly change AUC estimates. For example, when predicting BMI >30 kg/m^2, an AUC=0.588 was reported when using molecularly derived ancestry versus an AUC=0.586 when using self-identified ancestry in the model (full data not shown).

DISCUSSION

In this paper, we have constructed a GRSS comprised of 56 common polygenic variants and shown its association with BMI in 2,653 European-Americans and 973 African-Americans from the MGS-C sample. The GRSS was highly associated with BMI (p-value = 3.19×10^{-6}) and accounted for 0.66% of phenotypic variance in BMI. The association of the GRSS with BMI was comparable to sex, a known factor to influence body composition. The average effect of carrying 10 risk variants was an increase in BMI of 1.1 kg/m^2. This corresponds to a weight increase in an average male (5 feet 9 inches, 180 pounds) of 8 pounds and an average female (5 feet 4 inches, 155 pounds) of 7 pounds. Further, we have shown the association of the GRSS with BMI was not the

result of the few most significant SNPs but rather the aggregate of many SNPs of small effect. These results are consistent with the common disease common variants hypothesis indicating genetic variants common in the population with small effects contributes to the heritability of common traits and diseases.

ROC curves and the corresponding AUC estimates indicated statistical discriminative ability to predict obesity (BMI >30 kg/m^2, AUC=0.588, 95% CI= [0.567, 0.610]). AUC estimates were similar to those found in previous studies. For example, Renstrom *et al.* (140) used a genetic score of 9 SNPs and reported an AUC estimate of 0.575 in a sample of 353 obese and 1,370 normal weight diabetic and non-diabetic northern Swedes. Additionally, a study by Cheung *et al*. (2010) estimated an AUC of 0.582 with a genetic score including 13 SNPs in a Chinese sample of 470 obese cases and 700 normal weight controls. Although these AUC estimates were statistically significant, they were below 0.8, the threshold used in clinical practice for screening. In the MGS-C sample, however, the ability to predict morbid obesity (class III) was notably better and approached clinical criteria for a screening test (AUC=0.697, 95% CI= [0.663, 0.731]).

In the MGS-C sample, 4.1% of the phenotypic variance in BMI was accounted for using a model including sex, ancestry based on molecular derived principal components, age, and a GRSS comprised of 56 SNPs. Despite high heritability of BMI, much variance in BMI remains unaccounted for. Based on the progress in identifying loci influencing height, it is likely that a considerable portion of the 'missing heritability' resides in unidentified variants yet to be discovered by larger sample sizes (157). Large-scale international collaborative groups will be required to identify these additional variants with similar and smaller effect sizes.

Additionally, predictive models have yet to include other sources of variation known or hypothesized to influence BMI such as rare variants, gene-gene (GxG) or gene-environment (GxE) interactions, copy number variation, and epigenetic effects. For instance, rare variants which were not included in the current genetic risk profiles are likely to contribute to BMI heritability. For example, in a study by Blakemore *et al*.(2009), a rare variant in the visfatin gene was associated (p-value=8.0×10^{-5}, minor allele frequency 1.6% in control and 0.4% in obese subjects) with reduced risk for obesity. There is also evidence to support the influence of copy number variation (CNV) on BMI. In the Willer *et al*.(2009) meta-analysis, when examining CNV by SNP-CNV linkage disequilibrium, they found 10-kb and 45-kb deletion polymorphisms upstream of NEGR1 with the 45-kb deletion flanked by their two most associated BMI SNPs. The recent advent of SNP arrays designed for CNV detection may reveal additional genetic associations with BMI. Epigenetic variation, although more widely researched in syndromic obesity such as Prader-Willi, may also be linked to common obesity. Finally, GxG interactions have yet to be included in risk prediction of body composition. Twin studies support the role of non-additive genetic effects although most study designs have limited ability to detect them (35, 158).

Since obesity has increased dramatically while the genome has arguably remained stable, future research needs to address moderation effects of the environment. Known obesogenic factors such as physical activity and food intake have been shown to account for a significant portion of the variance in BMI with estimates ranging 5-10% (159-162). Additionally, research is beginning to elucidate GxE affecting BMI (163-167). At least two genes included in the current GRSS show evidence for GxE effects. For example,

Rampersaud *et al*.(2008), in a study of 704 Old Order Amish, found the effects of FTO variants associated with elevated body weight were attenuated in subjects with higher physical activity levels. Additionally, interactions between MCR4 and dietary intake and selection have been shown in model organisms (168-171). For example, mice when given a 3-choice diet and administered a melanocortin agonist preferentially decreased fat consumption (172). Further, variation in human MCR4 has been associated with binge eating (173-175) and with higher total energy intake and selection of foods high in dietary fat (165). BMI prediction models will benefit from incorporating known obesogenic environmental variables such as physical activity and food selection and intake.

The purpose of this study was to test a GRSS as an alternative approach to replication of association of common polygenic variation with BMI. As hypothesized the MGS-C sample had limited power to replicate individual loci when employing genome-wide significant thresholds even though there was strong *a priori* evidence of these variants to influence BMI. However, by constructing a GRSS summarizing the total number of risk alleles, the aggregate risk was found to be highly significantly associated with BMI. This study provides evidence of the utility of GRSS as an alternative approach to replication of common polygenic variation in complex traits. Furthermore, the results from the AUC analysis demonstrate meaningful progress towards a screening test that perhaps if used in conjunction with known obesogenic predictors such as physical activity and food selection and intake could identify persons for early environmental or medical intervention to prevent morbid obesity and the associated negative health consequences.

TABLES AND FIGURES

Figure 1: Frequencies of genetic risk sum score

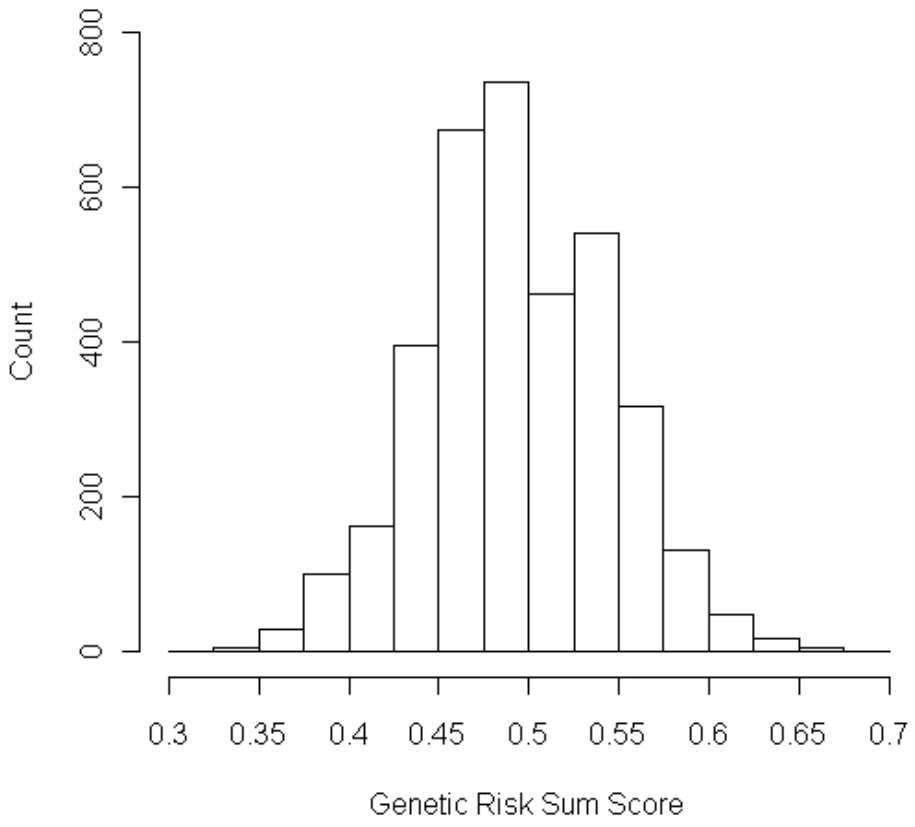
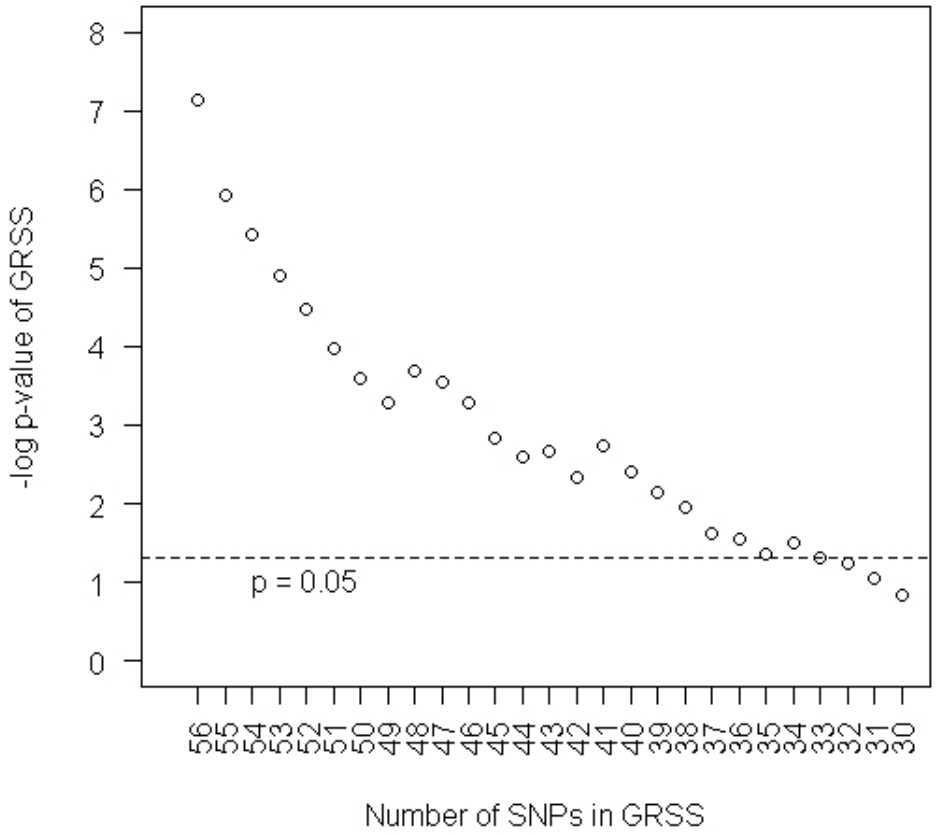


Figure 2: Number of SNPs in genetic risk sum score by –log significance of score



Note: GRSS = genetic risk sum score, -log = negative logarithm base 10, SNPs = single nucleotide polymorphisms.

Table 2: Descriptive statistics of MGS-C sample by race and sex

Group	*n*	*Mean*	*SD*
AA Males	381		
Age		46.59	13.39
BMI		29.62	5.95
AA Females	592		
Age		44.89	12.93
BMI		31.90	8.12
EA Males	1269		
Age		52.72	16.04
BMI		28.39	5.41
EA Females	1384		
Age		48.59	16.42
BMI		28.87	7.48

Note: AA = African-American, EA = European-American, Age = age in years, BMI = body mass index kg/m^2.

Table 3: Linear models predicting BMI in MGS-C sample

Model	*Estimate*	*SE*	*T*	*p*-value
Model 1: Covariates				
$F_{(6,3028)}$=19.18, *p-value* < 2.2E-16, *Adj. R*-squ = 0.0347				
Intercept	29.18	0.12	238.36	< 2E-16
PC1	94.78	11.99	7.90	3.8E-15
PC4	-49.19	19.05	-2.58	0.009
Sex	1.03	0.24	4.16	3.2E-05
Age	0.01	0.01	1.49	0.135
PC1*Sex	84.31	24.00	3.51	4.5E-04
PC4*Sex	-76.07	38.00	-2.00	0.045
Model 2: Covariates including GRSS				
$F_{(7,3027)}$=19.66, *p-value* < 2.2E-16, *Adj. R*-squ = 0.0413				
Intercept	29.18	0.12	239.17	< 2E-16
PC1	110.69	12.43	8.90	< 2E-16
PC4	-51.66	18.99	-2.72	0.006
Sex	1.03	0.24	4.20	2.7E-05
Age	0.01	0.01	1.50	0.132
PC1*Sex	85.57	23.91	3.57	3.5E-04
PC4*Sex	-74.42	37.87	-1.96	0.049
GRSS	11.41	2.44	4.66	3.2E-06

Note: BMI = body mass index kg/m^2, GRSS = genetic risk sum score, PC1 = principal components score distinguishes European from African ancestry, PC2 = principal components score distinguishes Eastern from Western European ancestry, Adj. R-squ = adjusted R-squared.

Table 4: Discriminative accuracy of genetic risk sum score and covariates predicting BMI category in the MGS-C sample

Group	*n* *(%)*	*AUC* *[CI]*	*Asy.* *Sig.*
Overweight	2157 (71.1)	0.613 [0.591,0.635]	1.21E-22
Obese 1	1139 (37.5)	0.588 [0.567,0.610]	3.11E-16
Obese 2	519 (17.1)	0.647 [0.621,0.673]	5.32E-26
Obese 3	232 (7.6)	0.697 [0.663,0.731]	1.75E-23

Note: BMI = body mass index kg/m^2, AUC = area under the receiver operator criteria curve, Asy. Sig. = asymptotic significance, Overweight = BMI >25 kg/m^2, Obese I = BMI >30 kg/m^2, Obese II = BMI >35 kg/m^2, Obese III = BMI >40 kg/m^2, predictors included in models: molecularly derived ancestry (principal components PC1 and PC4), sex, age, PC1 by sex and PC4 by sex interactions and genetic risk sum score.

SUPPLEMENTARY MATERIAL

Figure 3: MGS-C distribution of age in years

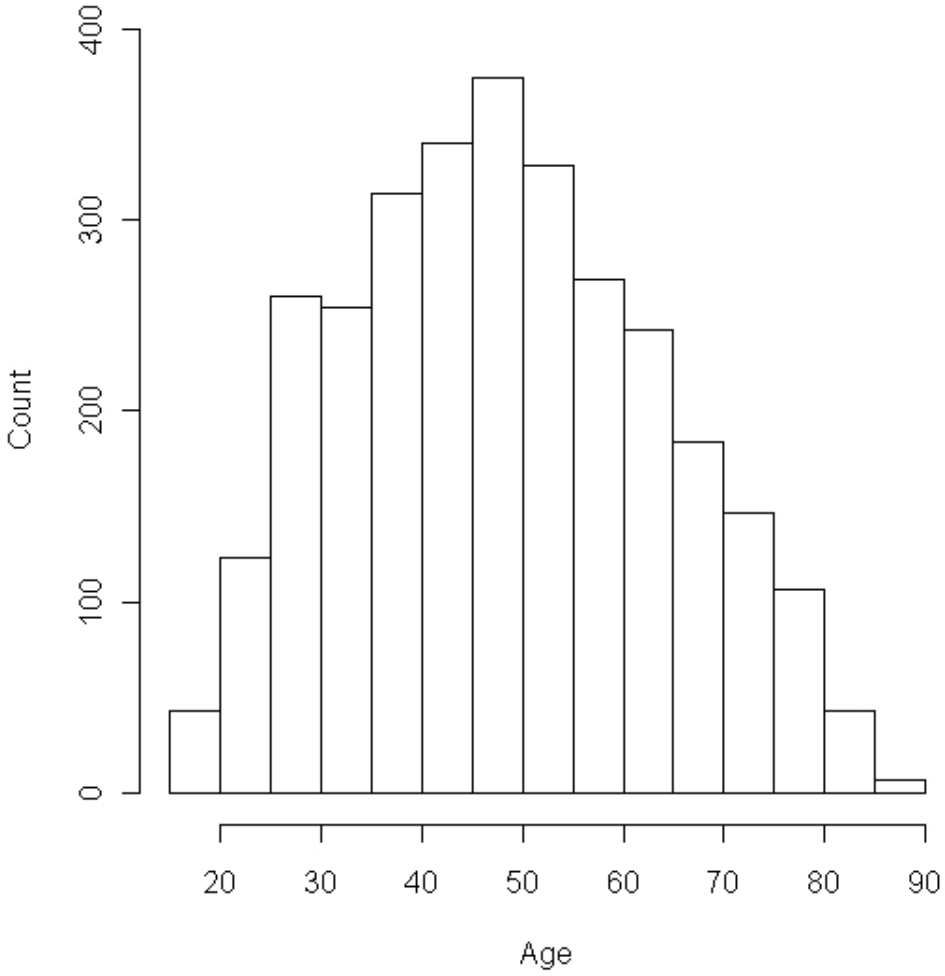
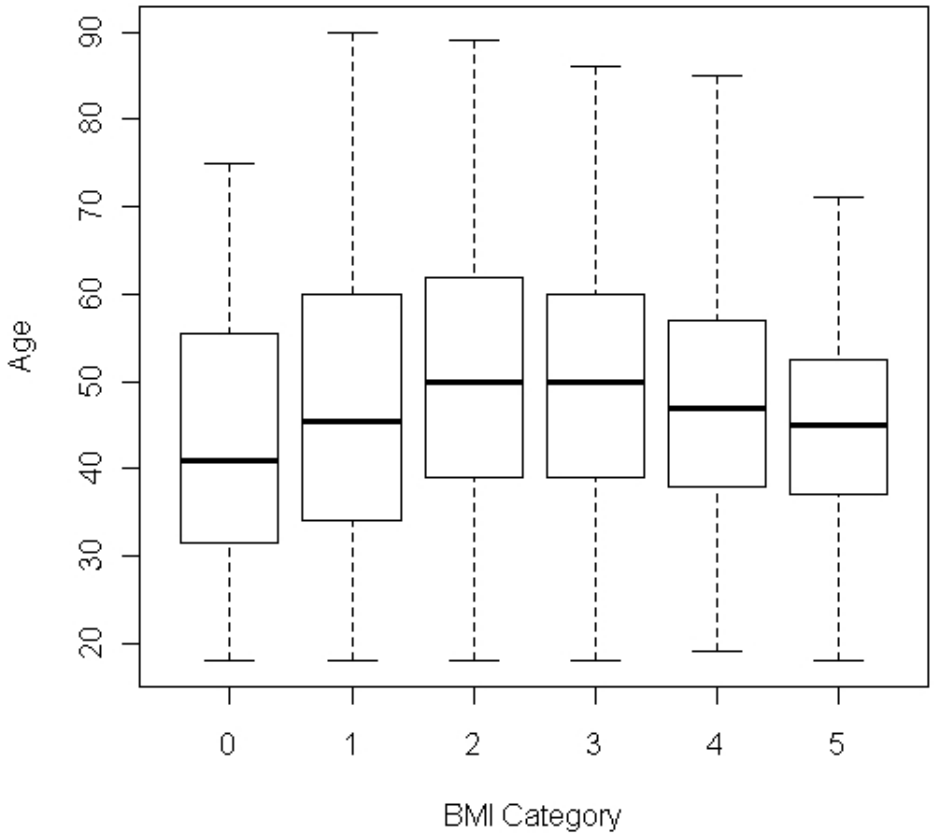


Figure 4: Mean age in years by BMI category



Note: BMI = body mass index kg/m^2, BMI category: 0 = underweight (BMI<18.5), 1 = normal-weight (18.5≤BMI <25), 2 = overweight (25≤BMI <30), 3 = obese class 1 (30≤BMI<35), 4 = obese class II (35≤BMI<40) and 5 = obese class III (40≤BMI), age = age in years, box plot is 95% .

CHR	SNP_meta	Allele	Var	Freq	Effect	SE	p-value	Paper	MGS_SNP	RSquared	Proxy	Allele_MGS	Variant_MGS	EA_freq_M	AA_freq_MG	NMISS_EA	BETA_EA	SE_EA	R2_EA	T_EA	P_EA	NMISS_AA	BETA_AA	SE_AA	R2_AA	T_AA	P_AA	Sum Score	BP(HG18)	Nearby Genes (BP)
1	rs3766431	T	C/T	0.446	3.87	0.90	1.6E10-05	Thorleifsson	rs3766430	0.97	1	C	T/C	0.4469	0.4147	2514	0.2149	0.1885	0.000517	1.14	0.2545	931	0.1278	0.3448	0.000148	0.3707	0.7109	1	54503239	LOC656495(-54) MRPL37(-46) SSBP3(0)
1	rs1776012	A	A/G	0.455	3.69	0.83	8.9E10-06	Thorleifsson	rs1776012	1.00	0	A	A/G	0.5008	0.352	2508	-0.2502	0.1894	0.000696	-1.321	0.1867	931	0.369	0.36	0.00113	1.025	0.3056	1	72490331	NEGR1(0)
1	rs9424977	T	C/T	0.458	3.71	0.81	4.6E10-06	Thorleifsson	rs9424977	1.00	0	T	C/T	0.5021	0.4778	2513	-0.2327	0.1888	0.000605	-1.232	0.2179	927	0.5135	0.3496	0.002328	1.469	0.1422	0	72497932	NEGR1(0)
1	rs2568958	A	A/G	0.581	4.66	0.87	9.9E10-08	Thorleifsson	Duplicate proxy SNP rs2815752																			0	72537704	NEGR1(-16)
1	rs2815752	A	A/G	0.62	0.15	0.04	9.30E-06	Willer	duplicate SNP of Thorleifsson et al. rs2815752																			0	72585028	NEGR1(-64)
1	rs3101336	G	A/G	0.582	4.56	0.86	1.1E10-07	Thorleifsson	rs2815752	1.00	1	A	G/A	0.6223	0.5714	2514	-0.3633	0.1924	0.001418	-1.889	0.05904	931	0.6699	0.3493	0.003943	1.918	0.05546	1	72585028	NEGR1(-64)
1	rs1973993	C	C/T	0.625	3.67	0.84	1.3E10-05	Thorleifsson	rs1973993	1.00	0	C	C/T	0.6033	0.24	2513	-0.1039	0.1907	0.000118	-0.5447	0.586	931	-0.1315	0.4007	0.000116	-0.328	0.743	1	96716582	
1	rs10783050	C	C/T	0.358	4.21	0.90	2.8E10-06	Thorleifsson	rs12562499	0.89	1	C	G/C	0.369	0.08428	2509	0.05769	0.1928	3.57E-05	0.2992	0.7648	925	0.1625	0.6354	7.08E-05	0.2557	0.7982	1	96720084	
1	rs10913469	C	C/T	0.2	4.29	0.93	4.2E10-06	Thorleifsson	rs543874	0.96	1	G	A/G	0.1957	0.2487	2511	0.4072	0.2341	0.001204	1.739	0.08211	931	0.4456	0.3991	0.00134	1.116	0.2645	1	176156103	LZTR2(8)
1	rs10910555	T	T/C	0.65				Willer	no proxy >.7																			0	230427413	
2	rs2867125	G	A/G	0.836	6.83	1.06	1.1E10-10	Thorleifsson	rs939583	1.00	1	T	C/T	0.8271	0.9029	2513	-0.398	0.2429	0.001068	-1.639	0.1014	931	-1.131	0.6014	0.003792	-1.88	0.06037	0	612531	TMEM18(47)
2	rs6548238	C	C/T	0.69	0.24	0.05	1.20E-06	Willer	rs6728726	0.925	1	C	T/C	0.8265	0.8926	2512	-0.4155	0.2422	0.001171	-1.716	0.08638	931	-1.038	0.5721	0.003529	-1.814	0.07001	0	613976	TMEM18(48)
2	rs4854344	T	G/T	0.839	6.33	1.00	2.9E10-10	Thorleifsson	rs13393304	1.00	1	G	A/G	0.8277	0.8885	2514	-0.4878	0.2433	0.001597	-2.004	0.04513	931	-1.178	0.5576	0.004781	-2.113	0.0349	0	627830	TMEM18(31)
2	rs7561317	G	A/G	0.837	6.42	1.01	2.4E10-10	Thorleifsson	rs10189761	1.00	1	A	T/A	0.828	0.7768	2512	-0.5	0.2434	0.001678	-2.054	0.04009	928	0.003127	0.4157	6.11E-08	0.007524	0.994	1	636364	TMEM18(23)
2	rs1439845	A	G/A	0.63				Willer	rs1439845	1	0	A	G/A	0.6129	0.8109	2513	0.2024	0.1904	0.00045	1.063	0.2878	931	0.2071	0.4431	0.000235	0.4675	0.6403	1	35264929	
2	rs10206343	T	T/C	0.41				Willer	no proxy >.7																			0	46019476	PRKCE(0)
3	rs7647305	C	C/T	0.772	4.66	1.00	3.1E10-06	Thorleifsson	rs1516728	0.86	1	T	T/A	0.7637	0.4697	2514	-0.1667	0.2184	0.000232	-0.7632	0.4454	931	-0.7229	0.3429	0.004762	-2.108	0.03527	1	187312585	ETV5(-6) DGKG(37)
4	rs752238	G	G/T	0.77				Willer	rs752238	1	0	G	G/T	0.7827	0.5936	2514	-0.2241	0.2298	0.000378	-0.975	0.3297	930	-0.09377	0.3516	7.66E-05	-0.2667	0.7898	1	44330777	ZNF595(43) ZNF718(43)
4	rs10938397	G	G/A	0.45	0.18	0.04	1.00E-05	Willer	rs12641981	1	1	T	C/T	0.434	0.2282	2513	0.7443	0.186	0.006335	4.001	6.49E-05	931	0.9161	0.4117	0.005301	2.225	0.02632	1	44874640	
4	rs2226814	T	T/G	0.11				Willer	no proxy >.7																			0	183378098	
5	rs17465346	T	T/A	0.95				Willer	rs10520924	0.92	1	A	A/G	0.9137	0.9676	2504	-0.03358	0.3282	4.18E-06	-0.1023	0.9185	930	-0.06217	0.9631	4.49E-06	-0.06455	0.9485	1	25464044	
5	rs467650	T	C/T	0.7	4.14	0.90	4.2E10-06	Thorleifsson	rs719246	0.812	1	T	C/T	0.6526	0.773	2506	-0.00852	0.1974	7.44E-07	-0.04316	0.9656	925	0.5619	0.4026	0.002107	1.396	0.1631	1	98085824	
5	rs2115172	A	A/G	0.565	2.45	0.74	9.70E-04	Thorleifsson	Duplicate proxy SNP rs10900767																			0	122713027	CCDC100(0)
5	rs10447440	G	A/G	0.57	2.51	0.75	8.40E-04	Thorleifsson	rs10900767	0.966	1	T	C/T	0.5678	0.5262	2514	-0.06072	0.1857	4.26E-05	-0.3269	0.7438	931	-0.2849	0.3481	0.00072	-0.8183	0.4134	1	122761602	CCDC100(-11)
6	rs2260000	T	C/T	0.672	2.75	0.78	4.60E-04	Thorleifsson	rs2736172	1.00	1	C	C/T	0.3787	0.08119	2511	0.1158	0.1931	0.000143	0.5995	0.5489	931	-0.6896	0.6403	0.001247	-1.077	0.2818	1	31698877	MCCD1(-83) BAT1(-82) U83(-86) U84(-81) ATP6V1G2(-76)
6	rs1977393	T	C/T	0.563	2.94	0.84	4.70E-04	Thorleifsson	Duplicate proxy SNP rs2736172																			0	31718508	
6	rs2844479	T	G/T	0.676	3.64	0.91	6.1E10-05	Thorleifsson	rs805297	0.83	1	C	C/A	0.682	0.944	2509	0.1404	0.2016	0.000193	0.6965	0.4862	931	-0.2322	0.7531	0.000102	-0.3084	0.7579	1	31730585	
6	rs12210863	G	G/A	0.86				Willer	rs6912289	0.818	1	G	G/A	0.8527	0.8493	2512	-0.5714	0.2617	0.001895	-2.183	0.02911	930	-0.08127	0.4653	3.29E-05	-0.1747	0.8614	1	40187497	
6	rs9511147	T	T/C	0.39				Willer	rs12055715	0.841	1	T	C/T	0.5051	0.222	2513	0.0209	0.1879	4.93E-06	0.1112	0.9114	931	-0.103	0.4145	6.65E-05	-0.2485	0.8038	1	46416395	DSCR1L1(-14)
6	rs9907460	A	A/G	0.39				Willer	rs6907460	1	0	A	A/G	0.4091	0.4655	2513	0.2786	0.1866	0.000887	1.493	0.1356	930	0.07249	0.3482	4.67E-05	0.2082	0.8351	1	54986006	FAM83B(-74)
6	rs9455181	T	T/C	0.23				Willer	rs10945231	0.725	1	G	A/G	0.1585	0.07534	2508	0.08724	0.2547	4.68E-05	0.3425	0.732	927	0.588	0.6496	0.000885	0.9052	0.3656	1	71354484	KIAA1411(-28) C6orf57(0) SMAP1(79)
7	rs16873846	T	T/C	0.95				Willer	no proxy >.7																			0	25348456	
7	rs11976955	C	C/G	0.69				Willer	rs7738134	0.878	1	G	C/G	0.6547	0.7383	2488	-0.03237	0.1962	1.09E-05	-0.1649	0.869	923	-0.01832	0.3958	2.33E-06	-0.04628	0.9631	1	130118115	TSGA13(-97) KLF14(-48) hsa-mir-29a(93) hsa-mir-29b-1(94)
8	rs17069257	C	C/T	0.133	3.91	0.95	3.6E10-05	Thorleifsson	rs7015381	0.94	1	C	C/T	0.152	0.1912	2512	0.1368	0.2558	0.000114	0.5349	0.5928	931	-0.2517	0.438	0.000355	-0.5747	0.5657	1	4032708	CSMD1(0)
8	rs11773921	A	A/T	0.14				Willer	rs4128268	0.774	1	A	G/A	0.1891	0.3022	2496	-0.2621	0.239	0.000482	-1.096	0.2731	921	0.1285	0.38	0.000124	0.3381	0.7354	1	84897493	
9	rs4743120	T	T/C	0.35				Willer	rs4743123	1	1	T	T/C	0.3622	0.1235	2513	0.3937	0.1879	0.001746	2.096	0.03622	930	0.4288	0.5234	0.000723	0.8194	0.4128	1	97440063	
9	rs6477693	C	A/C	0.308	3.13	0.82	1.50E-04	Thorleifsson	rs6477693	1.00	0	C	A/C	0.2829	0.3124	2503	-0.1405	0.2053	0.000187	-0.6847	0.4936	923	0.3535	0.3719	0.00098	0.9506	0.3421	1	110958186	C9orf4(0) EPB41L4B(83)
9	rs867559	G	A/G	0.217	5.07	1.13	7.0E10-06	Thorleifsson	rs867559	1.00	0	G	A/G	0.1907	0.3082	2509	0.04212	0.2371	1.26E-05	0.1776	0.859	928	0.194	0.3745	0.00029	0.5182	0.6045	1	128505146	LMX1B(-6)
10	rs4623795	G	G/C	0.84				Willer	rs4623795	1	0	C	G/C	0.8426	0.8791	2514	0.5537	0.2564	0.001853	2.159	0.03091	930	0.3067	0.5302	0.000361	0.5785	0.5631	1	7691915	ITIH1(0) ITIH5(0) ITIH2(93)
11	rs10769908	C	C/T	0.87				Willer	rs10769908	1	0	C	C/T	0.4992	0.3607	2513	0.3109	0.1844	0.001131	1.686	0.09188	931	0.1327	0.3555	0.00015	0.3733	0.709	1	8440665	STK33(0)
11	rs7481311	T	C/T	0.237	5.01	1.10	4.9E10-06	Thorleifsson	rs7481311	1.00	0	T	C/T	0.2324	0.2991	2513	0.7437	0.2194	0.004556	3.39	0.000709	931	0.3669	0.384	0.000981	0.9553	0.3397	1	27539705	LGR4(-89) LIN7C(-54) BDNP(96) BDNF(96)
11	rs4074134	G	A/G	0.844	5.77	1.21	1.8E10-06	Thorleifsson	rs879048	1.00	1	A	C/A	0.8011	0.6269	2502	-0.6444	0.2343	0.003016	-2.75	0.006	931	0.04437	0.3508	1.72E-05	0.1265	0.8994	0	27595510	BDNF(40) BDNF(40)
11	rs4923461	A	A/G	0.845	5.89	1.21	1.1E10-06	Thorleifsson	Duplicate proxy SNP rs10501087																			0	27613486	BDNF(22) BDNF(22)
11	rs10501087	T	C/T	0.844	5.68	1.23	4.2E10-06	Thorleifsson	rs7124442	1.00	1	T	T/C	0.8036	0.9322	2514	-0.6516	0.2331	0.003101	-2.795	0.005227	931	-0.3414	0.6853	0.000267	-0.4982	0.6185	0	27626684	BDNF(9) BDNF(9)
11	rs925946	T	G/T	0.335	4.9	0.94	2.0E10-07	Thorleifsson	rs7124442	0.98	1	C	C/T	0.3178	0.4584	2514	0.4578	0.2016	0.002049	2.271	0.02322	931	0.3335	0.3443	0.001009	0.9687	0.3329	1	27633617	BDNF(2) BDNF(2)
11	rs6265	G	A/G	0.853	5.67	1.26	7.2E10-06	Thorleifsson	rs6265	1.00	0	G	A/G	0.8249	0.9589	2514	-0.739	0.2413	0.003719	-3.062	0.002221	931	-1.333	0.8787	0.00247	-1.517	0.1297	1	27636492	
11	rs10835211	A	A/G	0.264	4.11	0.95	1.5E10-05	Thorleifsson	rs11030102	0.88	1	C	C/G	0.2561	0.1044	2514	0.5047	0.2146	0.002197	2.352	0.01875	930	0.9735	0.5636	0.003204	1.727	0.08448	1	27638172	BDNF(-1) BDNF(0)
12	rs4752856	A	A/G	0.34	0.15	0.04	7.10E-06	Willer	rs4752856	1	0	A	A/G	0.35	0.1038	2514	0.02042	0.1943	4.40E-06	0.1051	0.9163	931	0.4113	0.5639	0.000572	0.7294	0.466	1	47604618	KBTBD4(-48) NDUFS3(-42) C1QTNF4(-35) MTCH2(0) AGBL2(33)
12	rs7138803	A	A/G	0.373	4.89	1.00	9.6E10-07	Thorleifsson	rs7138803	1.00	0	A	A/G	0.3849	0.1814	2512	0.1975	0.1912	0.000425	1.033	0.3018	931	0.3458	0.4367	0.000674	0.7917	0.4287	1	48533735	TEGT(-90) LOC283331(-48) LOC144233(-10) FAIM2(16) AQP2(97)
12	rs836964	T	C/T	0.893	6.24	1.32	2.2E10-06	Thorleifsson	rs417644	1.00	1	T	T/A	0.8825	0.8212	2511	0.1993	0.2884	0.00019	0.691	0.4896	931	-0.3262	0.445	0.000578	-0.733	0.4637	1	48563693	LOC283331(-78) LOC144233(-40) FAIM2(0) AQP2(67) AQP5(78)
12	rs275982	A	A/C	0.23				Willer	rs11173067	0.705	1	G	A/G	0.2831	0.2613	2514	-0.1118	0.2045	0.000119	-0.5467	0.5847	930	0.5373	0.3863	0.00208	1.391	0.1646	1	58203089	
12	rs925460	C	C/T	0.78				Willer	rs2017594	0.722	1	T	G/T	0.8063	0.6553	2510	-0.2212	0.2354	0.000352	-0.9396	0.3475	930	0.1535	0.3522	0.000205	0.4359	0.663	1	121197161	MLXIP(-2) IL31(-25) B3GNT4(57) DIABLO(61) DIABLO(61)
13	rs7336332	G	A/G	0.154	5.02	1.11	6.3E10-06	Thorleifsson	rs9581881	0.97	1	C	C/T	0.1363	0.02518	2514	0.02299	0.2735	2.81E-06	0.08403	0.933	931	2.083	1.128	0.003658	1.847	0.0651	1	27004637	MTIF3(-92) LNX2(-16) POLR1D(89) POLR1D(89)
15	rs2255042	C	C/T	0.92				Willer	rs2255042	1	3	G	G/A	0.8986	0.3582	2513	0.02247	0.3139	2.04E-06	0.07158	0.9429	931	0.324	0.3546	0.000898	0.9137	0.3611	1	41621610	TP53BP1(-49) MAP1A(-10) HISPPD2A(0) CKMT1B(52) STRC(57)
15	rs12324805	C	C/A	0.31				Willer	rs12324805	1	0	C	C/A	0.3189	0.1475	2514	-0.1611	0.1965	0.000267	-0.8195	0.4126	931	-0.6336	0.4887	0.001806	-1.296	0.1951	1	80139255	RKHD3(-14) EFTUD1(70)
15	rs7181095	T	T/C	0.61				Willer	no proxy >.7																			0	91375823	CHD2(-11) RGMA(-11)
15	rs8024593	G	G/A	0.06				Willer	rs8024593	1	0	G	G/A	0.1121	0.5764	2514	-0.2584	0.2916	0.000312	-0.8859	0.3757	927	-0.599	0.3438	0.003271	-1.742	0.08178	1	96988019	IGF1R(22)
16	rs8049439	C	C/T	0.437	3.92	0.87	6.0E10-06	Thorleifsson	rs8049439	1.00	0	C	C/T	0.3828	0.4445	2514	0.4022	0.1912	0.001759	2.104	0.03547	931	-0.1744	0.3392	0.000285	-0.5142	0.6072	0	28745016	EIF3S8(-90) ATXN2L(0) ATXN2L(0) ATXN2L(0) ATXN2L(0)
16	rs4788102	A	A/G	0.436	4.03	0.87	3.5E10-06	Thorleifsson	rs8062405	1.00	1	G	A/G	0.3756	0.2775	2514	0.4273	0.1909	0.001991	2.239	0.02525	931	0.3064	0.3724	0.000728	0.8226	0.411	1	28745407	EIF3S8(-91) ATXN2L(0) ATXN2L(0) ATXN2L(0) ATXN2L(0)
16	rs7498665	G	A/G	0.438	4.25	0.89	1.7E10-06	Thorleifsson	Duplicate proxy SNP rs8062405																			0	28790742	ATXN2L(-35) ATXN2L(-35) ATXN2L(-35) ATXN2L(-35) TUFM(-25)
16	rs7498665	G	G/A	0.36	0.17	0.04	5.40E-06	Willer	rs11646653	0.845	1	C	C/T	0.3181	0.2539	2495	0.2776	0.2007	0.000767	1.383	0.1667	929	0.5231	0.3896	0.001941	1.343	0.1797	1	28829650	ATXN2L(-74) ATXN2L(-74) ATXN2L(-74) ATXN2L(-74) TUFM(-64)
16	rs8499640	A	A/G	0.647	5.55	1.02	6.0E10-08	Thorleifsson	rs7203521	1.00	1	A	G/A	0.5861	0.6393	2514	-0.1845	0.1881	0.000383	-0.9809	0.3267	931	-0.5333	0.3586	0.002376	-1.487	0.1372	1	52326794	KIAA1005(-34) FTO(0)
16	rs7190492	G	A/G	0.621	6.85	0.97	2.0E10-12	Thorleifsson	rs12446228	1.00	1	G	A/G	0.6212	0.7603	2514	-0.196	0.19	0.000424	-1.032	0.3023	931	-0.4934	0.3993	0.001641	-1.236	0.2168	1	52357888	KIAA1005(-65) FTO(0)
16	rs8044769	C	C/T	0.518	6.85	0.85	7.9E10-16	Thorleifsson	rs8057044	0.93	1	T	A/G	0.4843	0.7014	2511	0.3088	0.1851	0.001108	1.668	0.09546	926	-0.399	0.379	0.001198	-1.053	0.2928	0	52370115	KIAA1005(-77) FTO(0)
16	rs8050136	A	A/C	0.408	8.28	0.82	4.4E10-24	Thorleifsson	rs8050136	1.00	0	A	A/C	0.4024	0.4306	2514	0.3966	0.1886	0.001758	2.103	0.03555	931	-0.07266	0.3463	4.74E-05	-0.2098	0.8338	1	52373776	KIAA1005(-81) FTO(0)
16	rs3751812	T	G/T	0.407	8.92	0.88	3.3E10-24	Thorleifsson	Duplicate proxy SNP rs8050136																			0	52375961	KIAA1005(-83) FTO(0)
16	rs9939609	A	A/T	0.41	0.29	0.04	6.30E-17	Willer	rs9939609	1	3	T	C/T	0.4071	0.4722	2509	0.3997	0.189	0.001781	2.115	0.03452	929	-0.1242	0.3418	0.000142	-0.3632	0.7165	1	52378028	KIAA1005(-85) FTO(0)
17	rs11078773	C	C/T	0.51				Willer	rs11078773	1	0	C	C/T	0.4869	0.7959	2504	-0.1721	0.1861	0.000342	-0.9251	0.355	923	1.238	0.4382	0.008589	2.825	0.004834	1	8845288	PIK3R5(-89) NTN1(21)
18	rs633265	A	A/C	0.464	3.14	0.80	8.9E10-05	Thorleifsson	rs633265	1.00	0	A	A/C	0.4455	0.7667	2513	0.1372	0.1864	0.000216	0.7359	0.4619	931	0.1498	0.399	0.000152	0.3753	0.7075	1	55982448	
18	rs1350341	T	C/T	0.461	3.14	0.79	6.4E10-05	Thorleifsson	rs1673518	1.00	1	C	A/G	0.4474	0.7575	2514	0.1052	0.186	0.000127	0.5657	0.5716	931	-0.1804	0.3961	0.000223	-0.4554	0.6489	0	55986345	
18	rs17782313	C	C/T	0.21	0.21	0.04	3.90E-07	Willer	rs17782313	1	3	G	A/G	0.2446	0.2828	2505	0.2296	0.2166	0.000449	1.06	0.2892	927	-0.04195	0.3796	1.32E-05	-0.1105	0.912	1	56002077	
18	rs12970134	A	A/G	0.299	4.23	0.90	2.6E10-06	Thorleifsson	rs12970134	1.00	0	A	A/G	0.2743	0.1334	2510	0.2877	0.2063	0.000775	1.395	0.1632	929	0.6061	0.5221	0.001452	1.161	0.246	1	56035730	
19	rs1423052	G	G/C	0.26				Willer	no proxy >.7																			0	8922559	MUC16(0)
19	rs29941	C	C/T	0.691	4.25	0.94	5.6E10-06	Thorleifsson	rs29942	1.00	1	G	A/G	0.6806	0.8474	2499	0.1437	0.199	0.000209	0.7221	0.4703	922	-0.6423	0.4999	0.001791	-1.285	0.1991	1	39001117	OR4F17(61)
19	rs11084753	G	G/A	0.67	0.22	0.04	2.60E-07	Willer	rs11084753	1	0	G	G/A	0.6749	0.6439	2514	0.1745	0.198	0.000309	0.8811	0.3784	931	-0.1057	0.3583	9.36E-05	-0.2949	0.7681	1	39013977	CHST8(-58) KCTD15(-19)
20	rs2145270	T	T/C	0.65				Willer	rs2145270	1	0	T	T/C	0.6328	0.501	2512	0.1383	0.1944	0.000202	0.7114	0.4769	929	-0.6793	0.3535	0.003967	-1.922	0.05497	1	6569685	
20	rs1076052	T	C/T	0.45				Willer	rs1076052	1	3	T	C/T	0.4224	0.2011	2472	-0.3151	0.1868	0.00115	-1.687	0.09181	915	-0.1178	0.4207	8.60E-05	-0.2801	0.7794	1	13235505	TASP1(83)
22	rs4823535	G	G/A	0.69				Willer	rs5768614	1	1	G	G/A	0.7045	0.7838	2505	0.1822	0.2016	0.000326	0.9037	0.3663	927	-0.1024	0.4184	6.48E-05	-0.2449	0.8066	1	47209058	

NOTE: EA = European-American, CHR = chromosome, SNP_meta = SNP listed in meta-analyses, Allele = BMI increasing allele, Var = variant, Freq = BMI increasing allele frequency,
Effect = Thorleifsson SNPs effect in percentage of the standard deviation & Willer beta per allele change in BMI, SE = standard error, Paper = indicates from which meta-analyses SNP was catalogued from,
MGS_SNP = SNP used in MGS-C analyses, Rsquared = pairwise linkage disequilibrium between SNP from meta-analyses and SNP used in MGS analysis,
Proxy = indicates if SNP used in MGS-C analyses was exact SNP from meta-analyses(0) or a proxy SNP(1), freq_MGS = allele frequency of MGS allele, NMISS = number of individuals with non-missing genotypes,
BETA = beta estimate, R2 = squared correlation coefficient between SNP and BMI, T = t-value, P = p-value, Sum Score = included in GRSS(1), not included in GRSS(0), BP(HG18) = base-pair position in build HG18,
Nearby GENES (BP) = lists nearby genes and distance from SNP in base-pairs

Table 6: Linear model predicting BMI including GRSS interactions with covariates

Model	*Estimate*	*SE*	*T*	*p*-value
Model : Covariates including GRSS				
$F_{(11,3023)}$ = 13.06, *p-value*< 2.2E-16, *Adj. R*-squ = 0.0418				
Intercept	29.22	0.13	227.85	< 2E-16
PC1	115.40	13.42	8.60	< 2E-16
PC4	-50.01	19.34	-2.58	0.009
Sex	1.03	0.25	4.17	3.1E-05
Age	0.01	0.01	1.58	0.113
PC1*Sex	98.06	24.86	3.94	8.1E-05
PC4*Sex	-76.75	38.06	-2.02	0.044
GRSS*PC1	164.20	264.30	0.62	0.534
GRSS*PC4	-105.40	372.60	-0.28	0.777
GRSS*Sex	7.69	4.99	1.53	0.123
GRSS*Age	-0.20	0.15	-1.32	0.186
GRSS	11.69	2.48	4.71	2.6E-06

Note: BMI = body mass index kg/m^2, GRSS = genetic risk sum score, PC1 = principal components score distinguishes European from African ancestry, PC2 = principal components score distinguishes Eastern from Western European ancestry, Adj. R-squ = adjusted R-squared.

Chapter 3: Association of common and rare variation influencing body mass index: A combined single nucleotide polymorphism and copy number variation analysis

Adapted from: Roseann E. Peterson, Hermine H. Maes, Peng Lin, John R. Kramer, Victor M. Hesselbrock, Lance O. Bauer, John I. Nurnberger, Jr., Howard J. Edenberg, Danielle M. Dick and Bradley T. Webb. On the association of common and rare variation influencing body mass index: A combined Single Nucleotide Polymorphism and Copy Number Variation analysis. *European Journal of Human Genetics* (Submitted).

ABSTRACT

As the architecture of complex traits incorporates a widening spectrum of genetic variation, analyses integrating common and rare variation are needed. Body mass index (BMI) represents a model trait, since common variation shows robust association but accounts for a fraction of the heritability. A combined analysis of single nucleotide polymorphisms (SNP) and copy number variation (CNV) was performed using 2,348 European and African-Americans from the Study of Addiction: Genetics and Environment. Genetic risk sum scores (GRSS) were constructed using 32 BMI-validated SNPs and aggregate-risk methods were compared: count versus weighted and proxy versus imputation. The weighted SNP-GRSS constructed from imputed probabilities of risk alleles performed best and was highly associated with BMI ($p=4.3x10^{-16}$) accounting for 3% of the phenotypic variance. In addition to BMI-validated SNPs, common and rare BMI/obesity-associated CNVs were identified from the literature. Of the 84 CNVs previously reported, only 21-kilobase deletions on 16p12.3 showed evidence for association with BMI (p=0.003, frequency=16.9%), with two CNVs nominally associated with moderate-obesity, 1p36.1 duplications (OR=3.1, p=0.009, frequency 1.2%) and 5q13.2 deletions (OR=1.5, p=0.048, frequency 7.7%). All other CNVs, individually and in aggregate, were not associated with BMI or obesity. The combined model, including covariates, SNP-GRSS, and 16p12.3 deletion accounted for 11.5% of phenotypic variance in BMI ($p=3.34x10^{-54}$) and area-under-the-curve (AUC) estimates significantly predicted obesity classification with maximum discriminative ability for morbid-obesity (AUC = 0.750). Results show that incorporating validated effect-sizes and allelic probabilities improve prediction algorithms. Although rare-CNVs did not account for significant phenotypic variation, results provide a framework for integrated analytic approaches.

INTRODUCTION

Obesity, defined clinically by a body mass index (BMI) greater than 30 kg/m^2, is a serious public health problem that occurs in over 1/3 of American adults (11, 12, 127, 176) and is associated with numerous medical conditions including cardiovascular disease (13), type II diabetes (14, 15), cancer (16) and is comorbid with multiple psychiatric disorders (17-21). Although nutritional intake and physical activity affect relative body weight, twin and family studies have consistently shown a significant genetic contribution to body composition with heritability estimates of 40 to 70% (34-36).

Genome-wide association studies (GWAS) have successfully identified single nucleotide polymorphisms (SNPs) that contribute to inter-individual variation in BMI and common obesity (177, 178). To date, there are 32 SNPs showing robustly replicated association with BMI; these individually have small effects ranging 0.06 to 0.39 kg/m^2 change in BMI per risk allele and in aggregate they account for a limited proportion of the phenotypic variance (~1.45%) (63). The variant with largest effect, 0.39, is located in the first intron of the *fat mass and obesity-associated* (*FTO*) gene; this effect size corresponds to a weight increase per each risk allele of 2.5 pounds for an individual 5 feet 7 inches. The frequencies of the 32 risk-alleles range from 4 to 87% in populations of European descent (63-65).

Current GWAS designs are limited to detecting trait or disease associations with common variation in accordance with the common disease-common variant (CDCV) hypothesis (73). Although the number of robustly associated SNPs is limited, using the Genome-wide Complex Trait Analysis (GCTA) approach which uses all genetic variation measured, accounted for 17% of the phenotypic variance in BMI (179). However, this still leaves substantial heritability unaccounted for and has lead to efforts to identify rare variants contributing to common disease. Given the heritability of BMI and the observation that common SNPs only account for a portion of the expected phenotypic variance, additional classes of genetic variants such as structural and lower frequency variation are likely to influence body composition. Indeed there is a growing list of rare copy number variants (CNV) associated with BMI and obesity (74-82).

As the architecture of common complex traits and diseases has been associated with a widening spectrum of genetic variation, analyses integrating common and rare variation are needed. BMI represents a model trait for this approach, since common variation shows robust association but accounts for a limited portion of the heritability. Additionally, an increasing number of reports implicate structural and rare variation in BMI, which may account for a portion of the 'missing heritability'. Therefore, this study constructed and tested an integrated model of common and rare variation associated with BMI and obesity in 2,348 Americans of European and African descent from the Study of Addiction: Genetics and Environment (SAGE). We catalogued genetic variants associated with BMI and obesity from the literature, including common SNPs and common and rare CNVs. Given modest effect-sizes of common variants influencing BMI; the power to detect statistically significant genome-wide associations is limited. Therefore, instead of testing individual loci sequentially, a genetic risk sum score (GRSS) summarizing the total number of risk variants using loci with strong *a priori* evidence

was constructed and tested. We constructed SNP-GRSSs using 32 validated BMI-SNPs by both count and weighted methods. Additionally, to compare and extend existing methods, SNP-GRSSs using imputed genotype probabilities were constructed. Previously we applied the count method to a separate sample (180) and are extending this work to test weighted scores as well as scores constructed from imputed genotypes. Furthermore, common BMI/obesity-associated CNVs were tested individually as well as in aggregate by count scores. Given the limited power to detect low frequency variants (181, 182), rare BMI/obesity-associated CNVs were tested as collections by CNV-GRSSs. Additionally, since rare CNV burden scores have been associated with obesity (74, 77), the genome-wide load of rare CNVs was tested. Integrated linear and logistic regression models incorporated the following predictors via a stepwise process: standard covariates, SNP-GRSS, BMI/obesity-associated common CNVs, common CNV-GRSSs, rare BMI/obesity-associated CNV-GRSSs and rare CNV genome-wide burden scores. Furthermore, to assess clinical utility, the best fitting models were tested for obesity risk prediction by plotting receiver operator criteria (ROC) curves.

PARTICIPANTS AND METHODS

Participants and phenotypes

Participants were from the Study of Addiction: Genes and Environment (SAGE) (183) which was one of eight Phase 1 studies in the Gene Environment Association (GENEVA) consortium (http://genevastudy.org/) (184). The SAGE sample was drawn from three contributing projects, which have been previously described in detail: the Collaborative Study on the Genetics of Alcoholism (COGA) (185, 186), the Collaborative Study on the Genetics of Nicotine Dependence (COGEND) (187) and the Family Study of Cocaine Dependence (FSCD). The FSCD sample did not have body composition variables available for analysis and was not included in this analysis. All SAGE participants provided written informed consent for genetic studies and agreed to share their DNA and phenotypic information for research purposes. All samples were de-identified and only subjects who consented to health research were included. The institutional review boards at all data collection sites granted approval for the use of the data.

Study variables were assessed by interview, using versions of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (188, 189). BMI was calculated from self-reported height and weight. Participants were removed from data analysis if they had missing data on either height or weight or if calculated BMI was less than 14.5 or greater than 60, as values not in this range were considered data entry errors. Clinical body weight categories were defined as overweight (BMI > 25), obese class I (BMI > 30), II (BMI > 35) and III (BMI > 40). Age was included as age at interview in years. AD was defined by the SSAGA according to DSM-IV criteria (190) and ND was defined as having a Fagerström Test for Nicotine Dependence score of 4 or greater as assessed from the SSAGA. Power calculations for genetic effects in the SAGE sample were computed using the software Quanto from variant frequency, effect-size, odds-ratio and percent variance accounted for by variants reported in original papers (191).

Genotyping

Samples were genotyped on the Illumina Human 1M beadchip at the Center for Inherited Diseases Research at Johns Hopkins University. Data cleaning procedures included detection of gender mis-annotation and chromosomal anomalies, cryptic relatedness, population structure, batch effects, and Mendelian and duplication error detection. Details of quality control procedures have been previously reported (183). To minimize effects of population stratification, principal components (PC) were constructed using EIGENSOFT 3.0 (192) and SMARTPCA (193). As recommended by *Patterson et al*., to avoid disruption of the eigenvalue structure, SNPs used to construct PC scores were pruned at $r^2 > 0.7$ to correct for dependence between markers (193) and also limited to autosomes. 577,039 SNPs were used to generate 10 PCs. To circumvent over-fitting, only PCs that were associated with BMI and indicative of ancestral background were used in subsequent analyses (192-194).

CNV calling

The Illumina 1M array has 1,072,820 probes (which includes 23,812 non-SNP "intensity-only" markers) that were used for CNV detection. Three widely-used programs were used for CNV calling: CNVPartition (Illumina StudioBead software), PennCNV (195), and QuantiSNP (196). Genomic waves were adjusted for CNVs called by PennCNV and QuantiSNP (197). Both PennCNV and QuantiSNP report a metric score for quality control purposes and as recommended by QuantiSNP documentation, CNV calls with a Log Bayes Factor (LBF) less than 10 were removed as well as poor quality samples based on quality control measures for CNV analysis as described in our previous work (198). CNV calls from the three programs were compared against each other and Combined CNV (CNVision.org) was used to integrate the calls from the programs (199). To increase the positive predicative rate (198), only CNVs that were called by at least two programs were analyzed. Given that calls in centromeric, telomeric and immunoglobin regions are prone to harbor false positives, CNV calls in those regions were removed from analyses (195, 200).

Selection of BMI/obesity-associated genetic variation

BMI SNPs were catalogued from a large-scale BMI meta-analyses by *Speliotes* and colleagues (63). In brief, the meta-analysis incorporated a two-stage approach in which GWAS was performed on 249,796 individuals from 46 studies in the first stage and association was performed in an additional 125,931 individuals from 42 studies in the second stage. The meta-analyses of both stages identified 32 SNPs reaching genome-wide significance ($p<5x10^{-8}$). Of the 32 validated BMI SNPs, 15 did not appear on the SAGE sample Illumina 1M array. Ungenotyped markers were ascertained by two approaches in order to compare methods: 1) imputation and 2) proxy SNPs. IMPUTE2 was used to phase the observed genotypes and impute unobserved genotypes (201, 202) using the 1000 Genomes phase 1 reference panel (release June 2011, b37) (203). The proxy method used the LD structure of the genome to identify highly correlated SNPs that appear on the array as proxies for the unobserved SNPs. For the 15 SNPs not present

on the array, proxies were identified using SNP Annotation and Proxy Search V2.1 (204) except for rs11847697, which did not have a highly correlated proxy SNP (r^2<0.7) on the Illumina 1M array and was therefore not included in SNP-GRSSs constructed by the proxy method. Haploview version 4.10 was used to verify phase and corresponding proxy alleles (148, 149). Table 11 details information on the 32 catalogued SNPs.

BMI and obesity associated CNVs were catalogued from research published between January 2008 and January 2012 via PubMed search. Case reports, typical of monogenic inheritance, were not included in the catalogue as the focus of the current study was on common complex obesity. There were 3 BMI (63, 64, 205) and 83 obesity-associated CNV regions identified from the literature (75-79, 81, 206-209). Table 12 details information on the 84 catalogued CNVs.

BMI SNP genetic risk sum scores

Common BMI-associated SNPs catalogued from the literature (n = 32) (63) were tested in aggregate by constructing GRSSs. There are primarily two methods for constructing genetic scores: count and weighted methods. The count method is the sum of the number of risk alleles, whereas the weighted method incorporates the sum of the number of risk alleles each weighted by its odds-ratio or effect size. In this study, the weighted scores were constructed from regression coefficients reported by *Speliotes et. al* (63). Count and weighted scores using the proxy method were calculated using the profile option in PLINK (150). If SNP information was missing in an individual then the scoring routine imputed expected values based on sample allele frequency. Count and weighted scores using imputed genotypes were constructed using R version 2.13.1 (210). Furthermore, to extend existing GRSS methodology, count and weighted scores were constructed using probabilities of imputed risk alleles (p) genotypes by the equation below. Count scores were calculated with β = 1 and weighted scores with β = effect-size of each risk allele (A) reported by *Speliotes et. al* (63) summed over the number of risk alleles in the score (n).

$$\left(\sum_{1}^{n}\beta\left[\left(2*p(AA)\right)+p(Aa)\right]\right)/n$$

CNV association

In the SAGE sample, CNVs were considered common if they had a frequency of 1% or greater and determined rare if the frequency was less than 1%. Common CNVs previously shown to be associated with BMI/obesity were tested individually and in aggregate by count scores. Rare CNVs were tested in aggregate by count scores constructed from CNVs 1) previously reported to be associated with BMI/obesity and 2) not previously associated with BMI/obesity (genome-wide burden of rare variants). CNVs previously reported to be associated with BMI/obesity were considered the same region in the SAGE sample if the CNV boundaries shared at least 40% overlap with the CNV boundaries reported in the literature. Additionally, since there is evidence that the positive predictive rate is increased for large CNVs, which is likely due to the increased number of probes in larger variants, common and rare scores were also constructed from

only CNVs larger than 100-kb to potentially reduce the number of false positive calls in the score (198).

Linear models

R (210) was used to fit linear and logistic regression models using established covariates for BMI including ancestrally informative PCs, sex and age. AD and ND were also included as covariates since SAGE is a sample selected for these traits. Predictors in linear models were included in a stepwise process and independent variables were centered to facilitate interpretation of effects. Interactions between all variables with significant main effects (n=8) were tested and included in the final model if the p-value of the interaction was less than the Bonferroni corrected significance level of 0.002.

Prediction of obesity

To test whether the combined model of common and rare variation had clinical utility for obesity risk prediction, diagnostic efficiency was assessed. One method is to graph a receiver operator criteria (ROC) curve, which is a plot of the true positive rate (sensitivity) against the false positive rate (1 - specificity) and calculate the corresponding area-under-the-curve (AUC). An AUC is the probability that the predictor is greater for cases than controls (153, 154). An AUC may range from 0.5, non-informative (no greater than chance), to a maximum of 1.0, perfect discrimination between cases and controls. Generally, an AUC of 0.80 is suitable for screening while 0.99 is acceptable for diagnosis (211). Binary logistic regression was used to calculate predicted probabilities of the models and was used as the predictor to generate ROC curves. Discriminative accuracy of the model to predict BMI category was estimated by calculating the AUC from ROC curves using SPSS Statistics version 19.0. The StAR software was used to test for statistical differences between ROC curves (212).

RESULTS

Phenotypic detail

Complete data on height, weight, AD, ND, genotypes and CNVs were available for 1850 European-American and 498 African-American SAGE participants. Descriptive statistics for study variables are presented by sex in Table 7. The mean age of participants was 39.8 and ranged from 18 to 77. The average BMI of the sample was 27.5 kg/m^2, which is considered overweight, with 26.9% of the sample being obese (Table 10). There was a significant race by sex interaction with BMI (t-test=6.84, $p=1.01x10^{-11}$) indicating that females and African-Americans tended to have greater BMI. Males were more likely to be AD ($\chi^2=286.02$, $p=3.65x10^{-64}$) and ND ($\chi^2=9.36$, p=0.002). The age by AD interaction was also significant (t-test=-3.11, p=0.002) indicating that older subjects were less likely to be AD. Additional sample characteristics have been previously reported (183).

BMI SNP-GRSS

The mean number of BMI risk alleles per person for the 32 validated SNPs was 28.5 (SD=3.4) with a range from 18 to 39. The frequencies and distribution are shown in Figure 5. Power analyses calculated for the SAGE sample indicated 80% power to detect only one of the 32 BMI-validated variants; rs1558902 in *FTO* (Table 11) and a sample size of 177,492 would be needed to detect the smallest of the BMI-SNP effects. Indeed only two of the 32 BMI-SNPs were significantly associated with BMI in the SAGE sample after correction for multiple testing which included SNPs in or near *FTO* and *BDNF*. However, the sample size of SAGE has 99% power to detect the 32 variants in aggregate (GRSS), based upon effect-sizes reported in *Speliotes et al.* 2010 (63). Associations of the SNP-GRSSs with BMI are displayed in Table 8 and were highly significantly associated with BMI ($p < 1.11 \times 10^{-12}$). To compare common methods for computing SNP-GRSSs, as well as extend existing approaches, six GRSSs were constructed: 1) proxy SNP score by count and 2) by weighted method, 3) imputed SNP score by count and 4) by weighted method and 5) imputed probability of risk allele score by count and 6) by weighted method (see METHODS section). In general, the SNP-GRSSs constructed by weighted methods performed better than count methods ($z > 7.3$, $p < 0.0001$) and increased the percent of variance accounted for by 0.5-0.9%. Additionally, SNP-GRSSs that were constructed from imputed genotype probabilities performed better than scores constructed by the proxy method ($z > 3.2$, $p < 0.001$) and increased the percent of variance accounted for by 0.1-0.4%. The SNP-GRSS constructed from weighted imputed allelic probabilities performed the best and accounted for 3% of the phenotypic variance in BMI.

CNV association

Eighty-four BMI/obesity-associated CNVs were catalogued from the literature and tested for association with BMI and obesity in the SAGE sample. Detailed information may be found in Table 12. Of the reported CNVs in the literature, only 11 had sufficient information on frequency and effect-size/OR for power calculations and only 2 of these had 80% power to be detected in the SAGE sample. Power calculations for CNV aggregate risk scores were not performed because most of the variants reported in the literature did not cite corresponding effect-sizes or ORs. Of the 84 CNVs catalogued from the literature, 46 were called in the SAGE sample; 21 of these were common, including 17 deletions and 4 duplications, and 25 were rare, including 10 deletions and 15 duplications. Of the common CNVs, only a 21-kb deletion on 16p12.3 showed evidence for association with BMI (β=-0.057, p=0.003, frequency=16.9%). This CNV was also nominally associated with obese class I (OR=0.743, p=0.022) and II (OR=0.630, p=0.020). Additionally, two common CNVs were nominally associated with moderate-obesity (obese class II BMI > 35) in the expected direction. The first was a duplication on 1p36.1 (OR=3.1, p=0.009, frequency 1.2%) which ranged in length from 49.3 to 150.8 kb with a median value of 66.4 kb. The second was a large deletion on 5q13.2 (OR=1.5, p=0.048, frequency 7.7%) and ranged in length from 577.5 to 2238 kb with a median value of 1635 kb. CNV-GRSSs were constructed separately for common and rare variants. Also, deletions and duplications were tested both together and separately as well as limited to large CNVs over 100 kb. None of the CNV-GRSSs, common or rare, were

significantly associated with BMI or obesity in the SAGE sample. Descriptive statistics as well as association results for CNV-GRSSs are presented in Table 13.

Linear models

Results from linear regression analyses are displayed in Table 9. Ancestry was accounted for by three principal components PC1, PC4 and PC8 with PC1 distinguishing between European and African ancestries. PC1 and PC8 were associated with BMI in the full sample and PC4 was associated with BMI in the European-American sample. The base model (Model 1), which included the standard covariates, PC1 by sex and age by AD interactions but no genetic component accounted for 8.3% of the variance in BMI. Model 2, which added the SNP-GRSS and the 21-kb deletion on 16p12.3 to the base model, fit significantly better [$F_{(3\ 2335)}$=27.9, p=9.79x^{-18}] and accounted for an additional 3.2% of phenotypic variance in BMI for a total of 11.5%. Interactions between the covariates and the SNP-GRSS were not significant except for sex, which suggested that the SNP-GRSS was equally associated with BMI in European and African-Americans and across age. No significant interactions between the covariates and the 21-kb deletion on 16p12.3 were found, which indicated that the CNV was comparably associated with BMI in males and females, European and African-Americans and across the age range observed in SAGE.

Obesity risk prediction

To test the discriminative accuracy of models to predict obesity classification, ROC curves were plotted and the corresponding AUCs were calculated. Three sets of nested models were tested: 1) covariates (molecularly derived ancestry, sex, age, ancestry by sex interaction), 2) covariates, SNP-GRSS and interaction with sex and 3) covariates, SNP-GRSS and three obesity-associated CNVs (the 21 kb deletion on 16p.12.3, the 66 kb duplication on 1p36.1, and the 1440 kb deletion on 5q13.2). Table 10 displays statistics from ROC curve analysis by BMI category. AUC estimates indicated the models significantly predicted overweight and obesity classification with maximum discriminative ability when employing model 3 to predict class III obesity (AUC = 0.750, 95% CI = [0.702, 0.7971]). Models that included genetic information had significantly greater AUCs than models only including covariates (Table 10).

DISCUSSION

We have constructed an integrated model of common and rare variation catalogued from the literature and demonstrated its association with BMI in 1850 European-American and 498 African-American SAGE participants. This is one of the first studies to incorporate both SNPs and CNVs into an integrated genetic analysis for BMI and obesity risk prediction. The best fitting model included standard covariates, SNP-GRSS and a 21-kb deletion on 16p12.3, and accounted for 11.5% of the phenotypic variance in BMI (p=3.34x10^{-54}).

The effects of BMI-associated SNPs were incorporated into the integrated model via an aggregate risk score. There were six SNP-GRSSs constructed from 32 validated

BMI-associated SNPs; count and weighted methods were compared. The weighted score constructed from imputed probabilities of risk alleles performed the best and was highly associated with BMI ($p=4.3x10^{-16}$), accounting for 3% of the phenotypic variance. Comparisons of SNP-GRSS methodology indicated the variance in BMI accounted for was increased by a third when weighted methods and imputed probabilities of risk alleles were incorporated. These findings highlight the value of large-scale meta-analysis validation efforts to characterized effect sizes for genetic variants. Our results suggest that incorporating well-characterized effect sizes into GRSSs as well as utilizing genotypic probabilities from imputation procedures may improve BMI prediction algorithms. Future research should test these methods for improved risk prediction in other complex traits and diseases.

Although there were 84 BMI/obesity-associated CNVs catalogued from the literature, only 46 were detected in SAGE and only one was significantly associated with BMI. Speliotes *et al*., first reported the deletion on 16p12.3 in a large-scale BMI meta-analysis because a common BMI-decreasing allele was highly correlated with the same 21 kb deletion (63). In the present study, the CNV was also moderately associated with obesity classes I and II. The closest gene to the deletion is *GPRC5B*, which codes for a G-protein coupled receptor (family C group 5 member B); this receptor is of unknown function, and resides 50 kb upstream of the CNV (RefSeq, July 2008). Our results provide further evidence of a common CNV associated with body composition and suggest follow-up functional studies are warranted to verify its relevance to mechanisms underlying body composition.

Additionally, two common CNVs were nominally associated with moderate obesity (obese class II BMI>35) in the expected direction. Both of these CNVs were originally reported to be associated with obesity in *Jarick et al.* (208). The first was a duplication on 1p36.1 and was originally reported to be associated with early-onset extreme obesity in 423 parent-offspring trios (208). The two closest genes were found within 50 kb downstream: *SYF2,* which codes for a nuclear protein which may be involved with pre-mRNA splicing, and *C1orf63,* an open reading frame (RefSeq, July 2008). The second common CNV of nominal significance with moderate obesity in the SAGE sample was a deletion on 5q13.2 (OR=1.5, p=0.048). This CNV was reported to be associated with early-onset extreme obesity in 423 parent-offspring trios and in a case-control sample of 453 extremely obese children/adolescents and 435 normal-weight and lean adults (208). This large deletion encompasses numerous genes, which are detailed in Supplemental Table 2.

With the exception of the three aforementioned CNVs, our results did not yield additional support for previously reported BMI/obesity-associated CNVs, either individually or in aggregate. There are several potential reasons for this. First, it is possible that the effect-size and frequency of variants were not large enough to be detected in the SAGE sample, even when examined in aggregate. Given the limited information on effect-sizes of the CNVs reported in the literature, assessing the power to detect these variants in the SAGE sample is not straightforward. Additionally, it is conceivable that the collections of CNVs examined here contained a greater number of false positives than true variants, which masked the potential for replication by risk scores. In fact, only 4 of the 84 CNVs identified from the literature have been associated with BMI/obesity in multiple studies (Supplemental Table 2). Large-scale BMI/obesity-

associated CNV meta-analyses are needed to validate variants and to characterize the magnitude of their effects. Another issue with CNV analysis is that the CNV calling methodologies from microarrays are limited, as most SNP-arrays were designed to measure common variation across the genome and not to primarily detect CNVs (213, 214). Furthermore, the resolution of arrays to call CNVs, as well as their boundaries, is limited by probe density and the use of different algorithms when applied to the same data may give inconsistent results (198, 215-221).

We also assessed whether the integrated models were clinically useful for obesity risk prediction. Our results indicated statistical discriminative ability to predict obesity classification from a model including standard covariates, SNP-GRSS and three obesity-associated CNVs (the 21 kb deletion on 16p.12.3, the 66 kb duplication on 1p36.1, and the 1440 kb deletion on 5q13.2). AUC estimates showed the models significantly predicted overweight and obesity classification with maximum discriminative ability when predicting class III obesity (AUC = 0.750, 95% CI = [0.702, 0.7971]). Previously, we had constructed a SNP-score by the count method comprised of 56 genome-wide significant as well as suggestive variants to predict obesity in the Molecular Genetics of Schizophrenia control sample and also found maximum discriminative ability when predicting class III obesity (AUC = 0.697, 95% CI = [0.663, 0.731]) (180). The present findings represent a 5% increase in the AUC although fewer markers were used but CNVs were also included. Other studies have used SNP-GRSS to predict obesity, which have incorporated 8-32 SNPs with corresponding AUC estimates ranging from 0.575 to 0.597 (63, 140, 142, 143). This study is one of the first to incorporate both SNP and CNV information into an integrated model predicting obesity classification. Although the AUC estimates were statistically significant, they were below 0.8, the threshold used in clinical practice for screening. However, the ability to predict morbid obesity (class III) approached clinical criteria for a screening test and performed better than previous genetic risk models predicting obesity.

There are several possible extensions of the work presented here. First, SAGE participants consisted of a selected sample for substance-use behaviors. It is possible that the findings reported here are not generalizable to the American population at large. Although we have included alcohol and nicotine dependence as covariates in all analyses, research has shown these phenotypes to have complex relationships with body composition (113, 222), and this may complicate interpretation to the general adult population. Additionally, despite incorporating aggregate risk scores, which analyze collections of variants simultaneously to increase power and reduce problems associated with multiple testing, it is possible that the SAGE sample may still lack adequate power to confirm associations in the literature. It is important to note, however, that inclusion of variants, which are not well validated in such scores, can reduce the efficiency of this method. It is likely that the strong association of the SNP-GRSSs and not the CNV-GRSSs with BMI is a result of the fact that the BMI-SNPs have been validated by large-scale meta-analysis while most of the CNVs have not. Therefore, future research should test for associations in both larger and population-based samples.

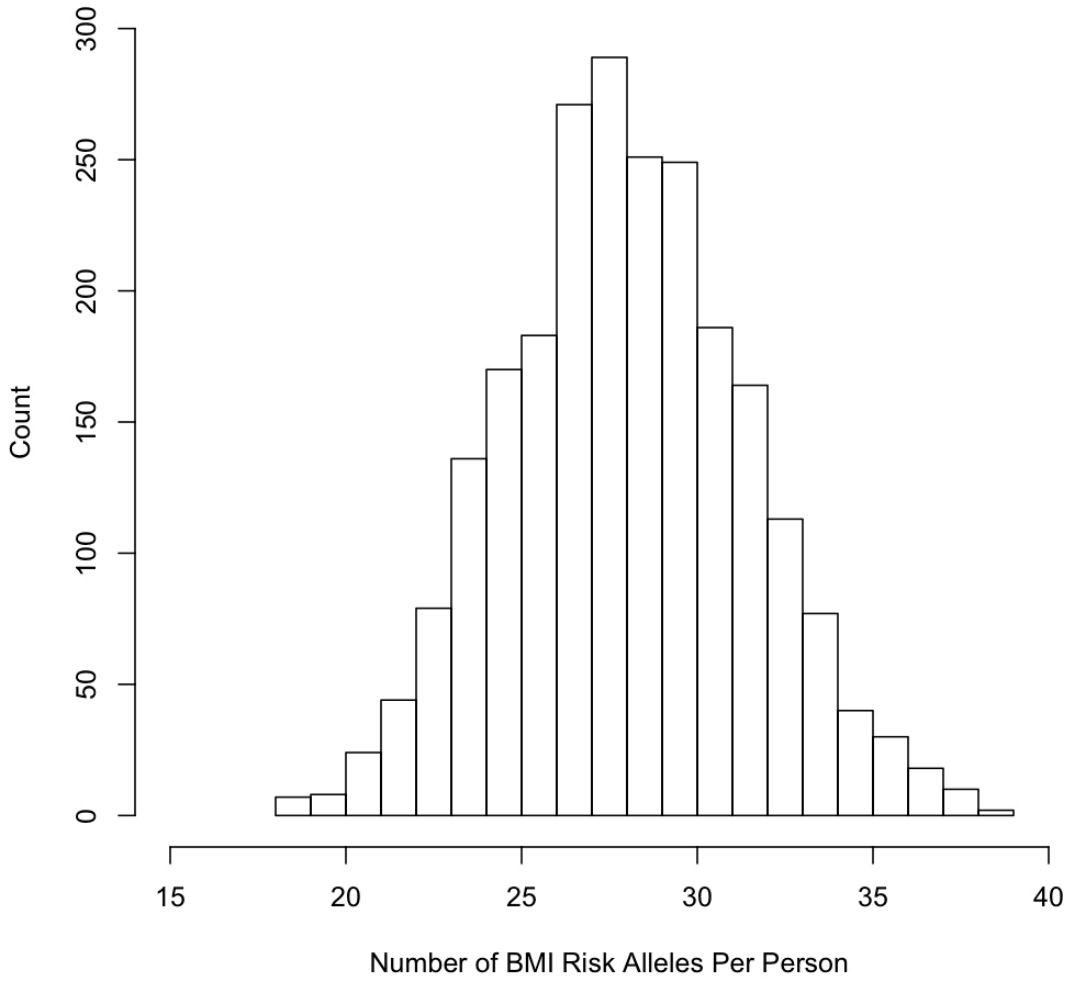
For many of the BMI/obesity-associated loci, it has yet to be determined if they do indeed represent the causative locus or if they are merely correlated with the causative variant. Fine mapping efforts are needed and will likely identify lower-frequency variants, which are typically not genotyped on commercial GWAS-arrays. As such, a

further extension of the work presented here is to include lower-frequency SNPs and INDELs identified by large-scale exome and genome sequencing efforts. Such studies are underway and include the UK10K project, a whole-genome sequencing study of 4,000 individuals and exome sequencing of an additional 6,000 individuals including 2,000 with extreme obesity phenotypes (83).

Furthermore, an important extension of an integrated model of BMI and obesity is to incorporate the moderating effects of the environment. Energy balance affects body composition, and research indicates that physical activity and food intake account for a significant portion of the variance in BMI, with estimates ranging from 5 to 10% (159-162). Additionally, at least two of the BMI-validated SNPs exhibit gene by environment interactions (GxE) (163, 165, 167, 173, 174, 223). For example, a large meta-analysis found that in physically active adults the effect of the *FTO* risk allele on obesity was attenuated by 27% (224). Given the considerable impact of the environment on body composition, future research needs to incorporate environmental variables into models of disease and risk prediction. Despite the potential limitations of the current study, this work provides a framework for integrating common and rare variation as both an alternative form of replication of genetic effects as well as for risk prediction of complex traits.

TABLES AND FIGURES

Figure 5: Frequency of BMI risk alleles per person (SAGE)



Note: BMI = body mass index kg/m^2.

Table 7: Descriptive statistics by sex in the SAGE sample

Group	***Males***		***Females***	
	Mean	***SD***	***Mean***	***SD***
Age	40.6	9.4	39.3	8.6
BMI	27.7	4.7	27.5	7.0
	N	***%***	***N***	***%***
	1011	43.7%	1337	56.3%
Obese	256	25.3%	376	28.1%
AD	672	66.5%	420	31.4%
ND	531	52.5%	617	46.1%

Note: Age = age at interview, BMI = body mass index kg/m^2, Obese = BMI > 30 kg/m^2, AD = alcohol dependence, ND = nicotine dependence.

Table 8: Comparison of GRSSs constructed by count and weighted methods

GRSS Method	*Mean (SD)*	*Estimate (SE)*	*T*	*p-value*	R^2
1. Proxy Count	0.450	15.99	7.18	9.07×10^{-13}	0.022
	(0.06)	(2.22)			
2. Proxy Weighted	0.063	126.22	8.56	2.05×10^{-17}	0.027
	(0.01)	(14.75)			
3. Imputed Count	0.447	16.28	7.16	1.11×10^{-12}	0.022
	(0.05)	(2.27)			
4. Imputed Weighted	0.062	128.75	8.51	2.94×10^{-17}	0.030
	(0.01)	(15.12)			
5. Imputed Probability Count	0.894	8.17	7.21	7.33×10^{-13}	0.022
	(0.11)	(1.13)			
6. Imputed Probability Weighted	0.124	64.42	8.54	2.43×10^{-17}	0.031
	(0.02)	(7.55)			

Note: GRSS = genetic risk sum score, Mean = mean score for GRSS, Estimate = regression coefficient for GRSS, Count = GRSS constructed from the summation of the number of risk alleles, Weighted = GRSS constructed from the number of risk alleles weighted by effect-sizes reported in *Speliotes et al.* 2010, SNP = single nucleotide polymorphism, Proxy = highly correlated substitute SNPs were used for variants not directly genotyped on the array, Imputed = genotypes were inferred from 1000 Genomes reference panel, Imputed probability = probability of genotypes inferred from 1000 Genomes reference panel.

Table 9: Linear models predicting BMI in the SAGE sample

Model	*Estimate*	*SE*	*T*	*p*-value
Model 1: Covariates [$F_{(9\ 2,338)} = 23.66$, p-value = 4.58×10^{-39}, $R^2 = 0.083$]				
Intercept	27.63	0.12	227.36	$< 2 \times 10^{-16}$
PC1	-98.82	8.67	-11.40	2.40×10^{-29}
PC4	10.54	7.63	1.38	0.167
PC8	-30.20	9.59	-3.15	0.002
Sex	-0.46	0.26	-1.75	0.081
Age	0.04	0.01	3.31	9.45×10^{-4}
AD	-0.20	0.07	-2.81	0.004
ND	-0.06	0.06	-0.91	0.361
PC1*Sex	-122.29	17.28	-7.08	1.92×10^{-12}
Age*AD	-0.02	0.01	-3.60	3.20×10^{-4}
Model 2: Covariates, GRSS & CNV [$F_{(12\ 2,335)} = 25.34$, p-value = 3.34×10^{-54}, $R^2 = 0.115$]				
Intercept	27.63	0.12	231.26	$< 2 \times 10^{-16}$
PC1	-110.22	8.72	-12.63	1.89×10^{-35}
PC4	10.14	7.50	1.35	0.176
PC8	-31.53	9.43	-3.34	8.36×10^{-4}
Sex	-0.43	0.26	-1.65	0.099
Age	0.04	0.01	3.35	8.15×10^{-4}
AD	-0.20	0.07	-2.81	0.005
ND	-0.07	0.06	-1.14	0.253
PC1*Sex	-131.38	17.26	-7.61	3.91×10^{-14}
Age*AD	-0.02	0.01	-3.41	6.59×10^{-4}
SNP-GRSS	62.44	7.62	8.19	4.30×10^{-16}
Sex*SNP-GRSS	44.37	15.19	2.92	0.003
Del 16p12.3	-0.57	0.32	-1.78	0.075

Note: BMI = body mass index kg/m^2, GRSS = genetic risk sum score, PC = principal component score reflecting ancestral background, Age = age at interview, AD = alcohol dependence, ND = nicotine dependence, CNV = copy number variation, Del = deletion.

Table 10: Discriminative accuracy of covariates, SNP-GRSS and CNV predicting BMI category in the SAGE sample

Model	AUC	95% CI	Asy. Sig.
Overweight: n = 1443 (61.4%)			
1. Covariates	0.679	[0.657,0.700]	$2.68x10^{-48}$
2. SNP-GRSS	0.692***	[0.671,0.714]	$9.23x10^{-56}$
3. CNV	0.694***	[0.672,0.715]	$1.27x10^{-56}$
Obese Class I: n = 632 (26.9%)			
1. Covariates	0.621	[0.594,0.647]	$2.74x10^{-19}$
2. SNP-GRSS	0.661***	[0.637,0.686]	$2.77x10^{-33}$
3. CNV	0.662***	[0.638,0.687]	$1.12x10^{-33}$
Obese Class II: n = 264 (11.2%)			
1. Covariates	0.648	[0.610,0.685]	$5.22x10^{-15}$
2. SNP-GRSS	0.681*	[0.646,0.716]	$6.97x10^{-22}$
3. CNV	0.690**	[0.656,0.725]	$5.58x10^{-24}$
Obese Class III: n = 106, (4.5%)			
1. Covariates	0.711	[0.660,0.762]	$1.97x10^{13}$
2. SNP-GRSS	0.741*	[0.692,0.790]	$4.81x10^{-17}$
3. CNV	0.750**	[0.702,0.797]	$3.15x10^{-18}$

Note: BMI = body mass index kg/m^2, SNP = single nucleotide polymorphism, SNP-GRSS = genetic risk sum score constructed from imputed probability of carrying 32 BMI-associated SNPs by the weighted method, CNV = copy number variation, AUC = area-under the receiver operator criteria curve, Asy. Sig. = asymptotic significance, Overweight = BMI > 25 kg/m^2, Obese I = BMI > 30 kg/m^2, Obese II = BMI > 35 kg/m^2, Obese III = BMI > 40 kg/m^2, Covariates = PC1, PC4, PC8, sex, age, AD, ND, PC1*sex, age*AD, PC = principal component score reflecting ancestral background, Age = age at interview, AD = alcohol dependence, ND = nicotine dependence, * = difference in AUC of the Model and Model 1 (Covariates) is $p < 0.05$, ** = difference in AUC of the Model and Model 1 (Covariates) is $p < 0.01$, *** = difference in AUC of the Model and Model 1 (Covariates) is $p < 0.001$.

FUNDING

This work was supported by the National Institute on Drug Abuse [DA26119]. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] [U01 HG004422]. SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center [U01 HG004446]. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA) [U10 AA008401], the Collaborative Genetic Study of Nicotine Dependence (COGEND) [P01 CA089392], and the Family Study of Cocaine Dependence (FSCD) [R01 DA013423, R01 DA019963]. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI [U01 HG004438], the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" [HHSN862200782096C].

ACKNOWLEDGEMENTS

We wish to thank the individuals who volunteered for the SAGE sample for their participation and the SARA Computing and Networking Services (www.sara.nl) for their support in using the Lisa Compute Cluster to generate principal component scores. Special thanks to Dr. T. Bernard Bigdeli, Ph.D., for guidance on imputation procedures and Dr. Charles O. Gardner, Ph.D., for statistical consultation.

SUPPLEMENTAL DATA

Supplementary material includes three tables detailing BMI/obesity-associated SNPs and CNVs catalogued from the literature and results of association analyses in the SAGE sample.

Table 11: 32 BMI SNPs

META-ANALYSIS Speliotes et al. 2010											SAGE SAMPLE																
SNP_meta	Chr	Position	effect_allele	other_allele	Freq_meta	Beta_meta	SE_meta	Explained Variance (%)	P_meta	Frq_EA	Beta_EA	SE_EA	T_EA	P_EA	BONF_EA	EMP_EA	Frq_AA	Beta_AA	SE_AA	T_AA	P_AA	BONF_AA	EMP_AA	Beta_meta_EA_AA	P_meta_EA_AA	Explained Variance (%)	Power:SAGE	Nearest Gene
rs2815752	1	72812440	A	G	0.61	0.13	0.02	0.04	1.61×10-22	0.628	0.062	0.02	2.896	**0.007**	0.227	**0.007**	0.546	0.021	0.04	0.491	0.623	1	0.620	0.053	**0.009**	0.240	0.183	NEGR1
rs11514175	1	74991644	A	G	0.43	0.07	0.02	0.02	8.16×10-14	0.424	0.044	0.02	1.892	0.059	1	0.058	0.652	-0.009	0.04	-0.205	0.838	1	0.836	0.032	0.115	0.009	0.105	TNNI3K
rs13555043	1	96344797	C	A	0.59	0.06	0.02	0.01	3.68×10-10	0.592	-0.007	0.02	-0.309	0.758	1	0.754	0.434	0.037	0.04	0.847	0.398	1	0.401	0.003	0.902	0.006	0.077	PTBP2
rs543874	1	177889480	G	A	0.19	0.22	0.03	0.07	3.56×10-23	0.196	0.052	0.02	2.260	**0.024**	0.766	**0.025**	0.267	0.112	0.04	2.595	**0.010**	0.312	**0.009**	0.065	**0.001**	0.455	0.250	SEC16B
rs2867125	2	622827	C	T	0.83	0.31	0.03	0.15	2.77×10-49	0.838	0.034	0.02	1.458	0.145	1	0.147	0.881	0.077	0.04	1.779	0.076	1	0.076	0.043	**0.034**	0.171	0.467	TMEM18
rs713586	2	25158008	C	T	0.47	0.14	0.02	0.06	6.17×10-22	0.475	0.032	0.02	1.388	0.165	1	0.170	0.845	-0.028	0.04	-0.631	0.528	1	0.527	0.019	0.347	0.049	0.221	RBJ
rs887912	2	59302877	T	C	0.29	0.1	0.02	0.03	1.79×10-12	0.296	0.034	0.02	1.456	0.146	1	0.143	0.091	0.047	0.04	1.064	0.288	1	0.283	0.036	0.075	0.114	0.134	FANCL
rs2890652	2	142559931	C	T	0.18	0.09	0.03	0.02	1.35×10-10	0.164	0.040	0.02	1.728	0.084	1	0.083	0.318	0.012	0.04	0.257	0.797	1	0.791	0.034	0.098	0.130	0.105	LRP1B
rs13078807	3	85884150	G	A	0.20	0.1	0.02	0.02	3.94×10-11	0.207	0.010	0.02	0.427	0.670	1	0.674	0.057	-0.040	0.04	-0.913	0.362	1	0.360	-0.001	0.962	<0.001	0.105	CADM2
rs9816226	3	185834499	T	A	0.82	0.14	0.03	0.03	1.69×10-18	0.819	0.060	0.02	2.620	**0.009**	0.283	**0.007**	0.817	0.029	0.04	0.658	0.511	1	0.505	0.054	**0.009**	0.226	0.134	ETV5
rs10938397	4	45182527	G	A	0.43	0.18	0.02	0.08	3.78×10-31	0.444	0.053	0.02	2.289	**0.022**	0.711	**0.021**	0.230	0.012	0.04	0.275	0.784	1	0.789	0.044	**0.031**	0.170	0.278	GNPDA2
rs13107325	4	103188709	T	C	0.07	0.19	0.04	0.03	1.50×10-13	0.069	0.036	0.02	1.566	0.118	1	0.117	0.016	0.092	0.04	2.111	**0.035**	1	**0.036**	0.048	**0.018**	0.153	0.134	SLC39A8
rs2112347	5	75015262	T	G	0.63	0.1	0.02	0.02	2.17×10-13	0.631	-0.010	0.02	-0.443	0.658	1	0.651	0.485	0.064	0.04	1.452	0.147	1	0.151	0.006	0.776	0.003	0.105	FLJ35779
rs4836133	5	124132103	A	C	0.48	0.07	0.02	0.01	1.97×10-9	0.535	0.022	0.02	0.941	0.347	1	0.341	0.816	0.051	0.04	1.155	0.249	1	0.243	0.028	0.170	0.044	0.077	ZNF608
rs206936	6	34302869	G	A	0.21	0.06	0.02	0.01	3.02×10-8	0.198	0.013	0.02	0.572	0.568	1	0.564	0.531	-0.075	0.04	-1.730	0.084	1	0.083	-0.006	0.758	0.016	0.077	NUDT3
rs987237	6	50803050	G	A	0.18	0.13	0.03	0.03	2.90×10-20	0.181	-0.013	0.02	-0.549	0.583	1	0.581	0.095	0.082	0.04	1.879	0.061	1	0.061	0.008	0.697	0.004	0.134	TFAP2B
rs4409776	9	22414339	G	A	0.31	0.11	0.02	0.02	2.65×10-13	0.310	-0.008	0.02	-0.354	0.724	1	0.727	0.158	0.082	0.04	1.885	0.060	1	0.057	0.012	0.565	0.008	0.105	LRRN6C
rs10968576	9	8604593	C	T	0.52	0.06	0.02	0.01	2.80×10-9	0.515	0.043	0.02	1.883	0.060	1	0.060	0.412	0.047	0.04	1.062	0.289	1	0.284	0.044	**0.031**	0.193	0.077	RP1-27A
rs4929949	11	27725986	A	T	0.78	0.19	0.03	0.07	4.69×10-26	0.796	0.075	0.02	3.278	**0.001**	**0.034**	**0.001**	0.950	0.012	0.04	0.269	0.788	1	0.782	0.062	**0.002**	0.386	0.250	BDNF
rs10767664	11	27655993	T	C	0.41	0.06	0.02	0.01	1.59×10-12	0.404	0.005	0.02	0.227	0.821	1	0.816	0.270	-0.087	0.04	-2.006	**0.045**	1	0.050	-0.015	0.459	0.012	0.077	MTCH2
rs3817334	11	47650993	T	C	0.38	0.12	0.02	0.04	1.82×10-17	0.376	0.026	0.02	1.127	0.260	1	0.258	0.172	0.005	0.04	0.118	0.906	1	0.907	0.021	0.293	0.038	0.183	FAIM2
rs7138803	12	50247468	A	G	0.24	0.09	0.03	0.02	9.48×10-10	0.220	-0.020	0.02	-0.841	0.401	1	0.399	0.218	-0.063	0.05	-1.385	0.167	1	0.173	-0.029	0.166	0.023	0.105	MTIF3
rs11847697	13	28020180	G	A	0.04	0.17	0.05	0.01	5.76×10-11	0.042	0.016	0.02	0.881	0.496	1	0.500	0.342	-0.022	0.05	-0.482	0.630	1	0.631	0.008	0.698	0.002	0.077	PRKD1
rs10150332	14	30515112	T	C	0.21	0.13	0.03	0.02	2.75×10-11	0.208	0.013	0.02	0.580	0.562	1	0.565	0.396	-0.043	0.04	-0.981	0.327	1	0.326	0.001	0.955	<0.001	0.105	NRXN3
rs2241423	15	68086838	G	A	0.78	0.13	0.02	0.03	1.19×10-18	0.768	0.004	0.02	0.193	0.847	1	0.849	0.629	0.015	0.04	0.350	0.726	1	0.722	0.007	0.738	0.001	0.134	MAP2K5
rs12444979	16	19933600	C	T	0.87	0.17	0.03	0.04	2.91×10-21	0.857	0.065	0.02	2.823	**0.005**	0.154	**0.005**	0.921	-0.018	0.04	-0.415	0.678	1	0.676	0.047	**0.021**	0.232	0.183	GPRC5B
rs7359397	16	28865659	T	C	0.40	0.15	0.02	0.05	1.88×10-20	0.382	0.030	0.02	1.318	0.188	1	0.188	0.059	0.035	0.04	0.776	0.438	1	0.441	0.031	0.127	0.082	0.192	SH2B1
rs1558802	16	53803574	A	T	0.42	0.39	0.02	0.34	4.8×10-120	0.420	0.086	0.02	3.753	**1.80E-04**	**0.006**	**3.00E-04**	0.097	0.067	0.04	1.529	0.127	1	0.130	0.082	**5.49E-05**	0.579	0.807	FTO
rs571312	18	57839769	A	C	0.24	0.23	0.03	0.1	6.43×10-42	0.236	0.039	0.02	1.686	0.092	1	0.089	0.348	-0.044	0.04	-1.020	0.308	1	0.308	0.021	0.312	0.032	0.335	MC4R
rs29941	19	34309532	G	A	0.67	0.06	0.02	<0.01	3.01×10-9	0.679	0.022	0.02	0.961	0.337	1	0.342	0.828	0.058	0.04	1.307	0.192	1	0.191	0.030	0.145	0.076	0.053	KCTD15
rs2287019	19	46202172	C	T	0.80	0.15	0.03	0.04	1.88×10-16	0.812	0.064	0.02	2.787	**0.005**	0.172	**0.006**	0.886	0.063	0.04	1.440	0.151	1	0.151	0.064	**0.002**	0.317	0.183	QPCTL
rs3810291	19	47569003	A	G	0.67	0.09	0.02	0.02	1.64×10-12	0.674	0.017	0.02	0.753	0.452	1	0.456	0.179	-0.013	0.04	-0.303	0.762	1	0.771	0.011	0.599	0.015	0.105	TMEM160

Note: BMI = body mass index, SNP = single nucleotide polymorphism, SAGE = Study of Addiction Genes and Environment, EA = European-American SAGE participants, AA = African-American SAGE participants, SNP_meta = SNP reported in large meta-analysis by Speliotes et al. 2010,
effect_allele = BMI increasing allele, Freq_meta = frequency of effect allele reported in large meta-analysis by Speliotes et al., Beta_meta = regression coefficient reported for effect allele in Speliotes et al., Frq = frequency of BMI effect allele, SE = standard error, T = student's t-statistic,
Bonf = Bonferonni corrected p-value for 32 tests, EMP = empirical p-value derived from 10,000 permutations, Beta_meta_EA_AA = regression coefficient for meta-analysis of EA and AA SAGE participants
Power = *post-hoc* power calculations for the SAGE sample derived from risk variant frequency, effect size and % variance accounted for as reported in Speliotes et al. 2010

Table 12: CNVs catalogued from the literature and frequency in the SAGE sub-sample

Supplementary Table 2: CNVs catalogued from the literature and frequency in the SAGE sub-sample

Region	Chr.	KB	Type	SAGE Count	SAGE Freq.	Com/Rare	EA_Count	EA_Freq	AA_Count	AA_Freq	Beta_BMI	P_BMI	OR_Ob1	P_Ob1	OR_Ob2	P_Ob2	OR_Ob3	P_Ob3	Reference	Phenotype	N Case / Control	Frequency Case / Control	OR (Effect Size)	Power SAGE	Build
chr6:52875284–52892054	6	16.8	Deletion																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr16:28291978-30099397	16	1807.4	Deletion																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr3:89245197–89343751	3	98.6	Duplication																Bochuva2010	obesity & developmental delay	300 / 7366	0.01 / 0	NR		NCBI 36.1
chr8:143067903–143688840	8	620.9	Duplication																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr10:432207–876909	10	444.7	Duplication	3	0.0012	Rare													Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0.0001	46.1*	0.143	NCBI 36.1
chr11:103489260–108818442	11	5329.2	Deletion																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr11:71980493–72106059	11	125.6	Duplication																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr15:28675248–30577016	15	1901.8	Duplication																Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0.0003	22.2*	0.287	NCBI 36.1
chr17:2218774–2259124	17	40.4	Duplication	3	0.0012	Rare													Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0.0003	22.2*	0.287	NCBI 36.1
chr22:49246176–49343366	22	103.2	Duplication	1	0.0004	Rare													Bochuva2010	obesity & developmental delay	300 / 7366	0.006 / 0	NR		NCBI 36.1
chr16:28700000–28900000	16	200.0	Deletion	1	0.0004	Rare													Bochuva2010, Bachmann-Gagescu2010	obesity & developmental delay	300 / 7366, 23084 / 7700	0.006 / 0.0003, 0.0013 / 0.00013	22.2*, 10.0*	0.287, 0.120	NCBI 36.1
chr16:29514353-30107356	16	593.0	Deletion	1	0.0004	Rare													Bochuva2010, Walters2010, Jacquemont20	obesity & developmental delay	00 / 7366, 7752 / 7434, 3544 / 5863	0.013 / 0.00054 , 0.0017 / 0.00013, 0.0073 / 0.0004	24.26*, 29.8, 18.25*	0.475, 0.160, 0.344	hg18
chr15:22140119–22151418	15	11.3	Duplication	81	0.0316	Common	69	0.035	12	0.022	-0.029	0.130	0.645	0.136	0.807	0.598	0.641	0.542	Chen2011	BMI	1000	0.0284	NR		NCBI 36.2
chr15:22275324–22278324	15	0.0	Duplication	59	0.0230	Common	54	0.027	5	0.009	-0.031	0.107	0.540	0.095	0.685	0.473	0.526	0.529	Chen2011	BMI	1000	0.0122	NR		NCBI 36.2
chr15:22289332-22300674	15	11.3	Duplication	50	0.0195	Common	48	0.024	2	0.004	-0.027	0.171	0.609	0.206	0.640	0.459	0.674	0.700	Chen2011	BMI	1000	0.0112	NR		NCBI 36.2
chr1:92164066-92168702	1	4.6	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr2:21337080-21342588	2	5.5	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr3:104059109-104092618	3	33.5	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr3:177280907-177303744	3	22.8	Duplication	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr4:35782722-35828015	4	45.3	Deletion	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0.00063	NR		hg18
chr5:53467427-53480255	5	12.8	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr5:77089051-77097628	5	37.6	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00013 / 0	NR		hg18
chr5:83835179-83874339	5	39.2	Deletion	3	0.0012	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.00046 / 0, AA = 0.00068 / 0	NR		hg18
chr5:90155080-90157383	5	2.2	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr6:31577721-31580099	6	13.0	Duplication	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr6:67361945-67395314	6	3.4	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0.0013	NR		hg18
chr7:113843696-113859679	7	16.0	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.002 / 0	NR		hg18
chr7:16885697-17020565	7	134.9	Duplication	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr7:20708193-20711088	7	2.9	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 /0	NR		hg18
chr7:8710455-8984816	7	274.4	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr9:115996089-116495996	9	499.9	Duplication	5	0.0020	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr10:118354808-118363319	10	8.5	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.00027 / 0, AA = 0 / 0.00063	NR		hg18
chr12:16141980-16149714	12	7.7	Deletion	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr12:1778115-1803814	12	25.7	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.0013 / 0.00063	NR		hg18
chr12:21446005-21459060	12	13.1	Duplication	8	0.0031	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr12:46232064-46330773	12	98.7	Duplication	4	0.0016	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr12:8322269-8401135	12	78.9	Duplication	5	0.0020	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00068 / 0	NR		hg18
chr13:42467085-42516530	13	49.4	Duplication	3	0.0012	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.00027 / 0, AA = 0.00068 / 0	NR		hg18
chr13:56772821-56786987	13	14.2	Deletion	55	0.0215	Common	7	0.004	47	0.085	-0.012	0.553	0.465	0.026	0.928	0.848	0.917	0.875	Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.00027 / 0, AA = 0.050 / 0.060	NR		hg18
chr13:61409439-61413267	13	3.8	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr16:79761583-79838201	16	76.6	Duplication	3	0.0012	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0037 / 0, AA = 0 / 0	NR		hg18
chr17:49444406-49449022	17	4.6	Duplication																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.00013 / 0	NR		hg18
chr19:10489548-10512171	19	22.6	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.0034 /0	NR		hg18
chr20:5111828-5137472	20	25.6	Deletion																Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0 / 0	NR		hg18
chr22:17405381-19770502	22	2365.1	Duplication	1	0.0004	Rare													Glessner2010	obesity	EA =1080 / 2500, AA = 1479 / 1575	EA = 0.0027 / 0, AA = 0.002 / 0.0025	NR		hg18
chr6:135605109-135818902	6	213.8	Duplication																Huang2012	overweight & BMI	70/74	0.243 / 0.054	5.62	0.999	hg19
chr16:29514353-30107356	16	593.0	Duplication																Jacquemont2011	underweight	58635	0.0005	8.30	0.293	hg18
chr1:147182865-147986383	1	775.5	Deletion	238	0.0930	Common	197	0.099	39	0.071	-0.017	0.380	0.930	0.648	1.273	0.240	0.869	0.697	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr1:25468522-25534812	1	66.3	Duplication	30	0.0117	Common	18	0.009	12	0.022	0.024	0.217	1.471	0.324	3.109	**0.009**	1.878	0.411	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr2:184502747-184510699	2	8.0	Deletion	7	0.0027	Rare													Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr2:41091935-41099404	2	7.5	Deletion	472	0.1844	Common	302	0.151	169	0.306	0.011	0.587	1.021	0.861	1.113	0.498	0.971	0.906	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr3:131245537-131290979	3	45.4	Deletion	19	0.0074	Rare													Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr3:133475451-133478387	3	2.9	Deletion	1	0.0004	Rare													Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr3:194360584-194365610	3	5.0	Duplication	2	0.0008	Rare													Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr3:89477270-89499280	3	22.0	Deletion	190	0.0742	Common	178	0.089	11	0.020	0.023	0.248	1.275	0.158	1.078	0.769	0.937	0.881	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr4:108285188-108293270	4	8.1	Deletion	116	0.0453	Common	111	0.056	5	0.009	-0.011	0.578	0.824	0.406	0.817	0.556	1.166	0.749	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr4:161282532-161289730	4	7.2	Deletion	120	0.0469	Common	112	0.056	7	0.013	-0.004	0.830	0.904	0.655	0.944	0.861	0.728	0.597	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr4:9832714-9844379	4	11.7	Duplication																Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr5:68903038-70343313	5	1440.3	Deletion	198	0.0773	Common	144	0.072	53	0.096	0.026	0.187	1.304	0.102	1.509	**0.048**	1.438	0.248	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr6:32560895-32638289	6	77.4	Deletion	1243	0.4855	Common	853	0.427	275	0.497	0.003	0.893	0.970	0.699	1.056	0.619	1.027	0.877	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr6:81340896-81346266	6	5.4	Duplication																Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr8:5586122-5591428	8	5.3	Deletion	306	0.1195	Common	281	0.141	24	0.043	-0.030	0.128	0.915	0.544	0.770	0.249	0.524	0.133	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr10:47011183-47145122	10	133.9	Deletion	68	0.0266	Common	24	0.012	44	0.080	-0.009	0.635	0.799	0.426	0.968	0.928	0.718	0.590	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr10:58186369-58198856	10	10.5	Deletion	76	0.0297	Common	62	0.031	14	0.025	-0.025	0.198	0.771	0.370	0.645	0.351	0.847	0.820	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr11:55130596-55210165	11	79.6	Deletion	1040	0.4063	Common	918	0.459	117	0.212	0.016	0.405	1.017	0.861	1.192	0.189	1.109	0.625	Jarick2011	obesity	1061 trios, 453 / 453	0.27 (parents)	1.19	0.647	hg18
chr12:9525125-9604954	12	79.8	Deletion	395	0.1543	Common	330	0.165	63	0.114	0.000	0.991	1.149	0.267	1.039	0.829	0.990	0.971	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr15:28280641-28609063	15	328.4	Deletion	65	0.0254	Common	37	0.019	28	0.051	-0.003	0.872	1.048	0.868	1.010	0.980	0.505	0.353	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr16:14796084-14987969	16	191.9	Duplication	15	0.0059	Rare													Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr17:41522227-42107492	17	585.3	Deletion	262	0.1023	Common	197	0.099	65	0.118	-0.008	0.674	0.988	0.935	0.695	0.123	0.781	0.492	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr19:20388034-20508242	19	120.2	Deletion	318	0.1242	Common	234	0.117	83	0.150	0.012	0.552	0.898	0.440	1.008	0.965	1.487	0.130	Jarick2011	obesity	1061 trios, 453 / 453		NR		hg18
chr1:65650000-65900000	1	250.0	Deletion																Jeron2010	T2D, obesity	1204		NR		
chr10:46338178-46812351	10	474.2	Deletion	24	0.0094	Rare													Shaw2009, Jarick2011	BMI, obesity	597, 1061trios & 453 / 453	0.05, NR	NR	0.891	hg18
chr16:19850972-19871980	16	21.0	Deletion	432	0.1690	Common	372	0.186	60	0.108	-0.057	**0.003**	0.743	**0.022**	0.630	**0.020**	0.810	0.478	Spellotes2010	BMI	125931	0.13	(0.17)	0.163	hg19
chr2:106245033-107807545	2	1562.5	Deletion																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr2:137328699-138602350	2	1273.7	Duplication																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr4:141598764-143656669	4	2057.9	Deletion																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr4:89822108-93149947	4	3327.8	Deletion																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr10:41756307-42943818	10	1187.5	Duplication																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr11:84695124-86095201	11	1400.1	Duplication																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr11:86463458-91574130	11	5110.7	Duplication																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr13:22153141-24201255	13	2048.1	Duplication																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr16:80739605-82222770	16	1483.2	Deletion																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr18:46630080-68330148	18	2167.1	Deletion																Wang2010	obesity	430 / 379	0.0023 / 0	NR		NCBI36
chr1:72765000-72810000	1	45.0	Deletion	13	0.0051	Rare													Wiler2009, Spellotes2010, Jarick2011	BMI, BMI, obesity	51112, 125931, 1061 trios	0.62, 0.61, 0.66 (parents)	(0.10), (0.13), 1.21	0.086, 0.163, 0.760	NCBI 35, hg19, hg18

Note: CNV = copy number variation, SAGE = Study of Addiction: Genes and Environment, EA = European-American, AA = African-American, Chr. = chromosome, KB = kilobase, Count = count of CNV, Freq = frequency of CNV,
Common = frequency of CNV greater than 1% in the SAGE sample, Rare = frequency of CNV less than 1% in the SAGE sample, BMI = body mass index, Beta_BMI = regression coefficient of CNV predicting BMI, OR = odds ratio,
Ob1 = obese class I (BMI > 30), Ob2 = obese class II (BMI > 35) , Ob3 = obese class III (BMI > 40), Reference = citation for CNV, Build = genome release reported in reference, Genes_hg19 = genes reported in CNV interval +/- 50kb referenced from genome build hg19.
NR = not reported or easily located in paper, Power = *post-hoc* power calcualtions for the SAGE sample based on characteristics of original paper, * = OR estimated by R.E.P. from sample sizes and frequencies reported in original paper.

Table 13: Common and rare CNV-GRSS

I. Genome-wide load of rare CNVs

	Median	Range	Beta BMI	P BMI	OR Ob1	P Ob1	OR Ob2	P Ob2	OR Ob3	P Ob3
Rare deletion genome-wide load	4	0-62	0.010	0.612	1.004	0.645	1.011	0.333	0.999	0.947
Large Rare +100 Kb deletion genome-wide load	0	0-40	0.014	0.473	1.009	0.598	1.022	0.305	1.026	0.426
Rare duplication genome-wide load	2	0-57	-0.015	0.444	1.01	0.533	1.006	0.805	0.963	0.442
Large Rare +100 Kb duplication genome-wide load	0	0-19	-0.018	0.363	1.041	0.241	0.949	0.363	0.891	0.268
Rare genome-wide load CNV	7	0-119	0.001	0.949	1.006	0.423	1.009	0.332	0.994	0.731
Large Rare +100 Kb genome-wide load CNV	1	0-54	0.003	0.878	1.014	0.334	1.009	0.637	1.008	0.801

II. CNV-GRSSs comprised of rare variants previously identified to be associated with BMI/obesity

	Count	Freq	Beta BMI	P BMI	OR Ob1	P Ob1	OR Ob2	P Ob2	OR Ob3	P Ob3
Rare obesity deletion	72	0.028	-0.017	0.37	0.657	0.177	1.087	0.827	1.505	0.435
Rare +100 Kb obesity deletion	46	0.017	-0.001	0.968	0.828	0.545	1.194	0.618	1.569	0.302
Rare obesity duplication	49	0.019	0.008	0.65	1.086	0.781	1.417	0.363	0.62	0.626
Rare +100 Kb obesity duplication	34	0.013	0.002	0.906	1.153	0.701	1.167	0.77	0.825	0.85
Rare obesity CNV	120	0.046	-0.008	0.698	0.838	0.413	1.235	0.441	1.143	0.769
Rare +100 Kb obesity CNV	80	0.031	0.001	0.968	0.941	0.798	1.187	0.564	1.381	0.437

III. CNV-GRSSs comprised of common variants previously identified to be associated with BMI/obesity

	Median	Range	Beta BMI	P BMI	OR Ob1	P Ob1	OR Ob2	P Ob2	OR Ob3	P Ob3
Common Obesity deletion	2	0-7	-0.015	0.449	0.969	0.359	1.008	0.874	0.982	0.807
Common Obesity duplication	0	0-3	-0.022	0.251	0.863	0.198	1.003	0.983	0.901	0.714
Common Obesity CNV	2	0-9	-0.021	0.28	0.959	0.202	1.007	0.876	0.975	0.733

Note: CNV = copy number variation, GRSS = genetic risk sum score, BMI = body mass index, Beta BMI = regression coefficient of GRSS predicting BMI, OR = odds ratio, Ob1 = obese class I (BMI > 30), Ob2 = obese class II (BMI > 35) , Ob3 = obese class III (BMI > 40), Kb = kilobase.

Chapter 4: Association of common polygenic variation with body mass index across adolescent development: A longitudinal twin study

Adapted from: On the association of common polygenic variation with body mass index across adolescent development: A longitudinal twin study. Roseann E. Peterson, Bradley T. Webb, Elizabeth C. Prom-Wormley, Judy L. Silberg, Lindon J. Eaves, and Hermine H. Maes. Presentation. The 42nd Annual Meeting of the Behavior Genetics Association. June 24th, 2012. Edinburgh, Scotland, UK.

ABSTRACT

A dramatic increase in the prevalence of obesity in developed countries and the numerous adverse consequences associated with elevated body weight in both children and adults highlight the necessity of research that aims to understand the genetic and environmental trajectories of relative body weight. Genome-wide association studies of body mass index (BMI) using large-scale adult samples have yielded 32 robustly associated genetic variants. Further research should address when during human development these variants begin to influence body weight. Therefore, we sought to utilize a developmental twin study design in order to determine the genetic and environmental architecture of BMI by variance components analysis and assess the effects of adult-validated BMI-SNPs across adolescence. Data analyses included 2,794 twin participants from the Virginia Twin Study of Adolescent Behavioral Development (ABD) ranging in age from 8 to 18 years old. BMI was calculated from weight and height for up to three waves of data collection. Variation in BMI at each age, as well as covariation across the age range was modeled using the independent pathway (IP) models which includes both genetic and environmental common and time-specific factors. BMI was found to be highly heritable, accounting for 74-91% of the variance over the course of adolescent development. Our best-fitting model indicated multiple genetic factors that contributed to BMI liability, including a genetic factor that loaded across development, a second common genetic factor that loaded later in adolescence and time-specific genetic factors important in mid-adolescence. Additionally, shared environmental effects were found to account for significant portions of the phenotypic variance (1-18%) for ages 11-16 in females and ages 8-14 in males. A unique environmental factor accounted for 2-13% of the phenotypic variance across development. To understand the importance of adult BMI-associated genetic variants across adolescent development, a genetic risk sum score (GRSS) was tested as an effect on latent genetic factors as well as on mean BMI. Preliminary results, assessed on a sub-sample of ABD twins, indicated that the GRSS was best modeled as an effect on mean BMI at each age group suggesting association across development with the magnitude of the effect differing at each time point considered. The GRSS accounted for 1-2.3% of the phenotypic variance in BMI across adolescence. These results, although preliminary, merit future research, which considers pubertal stage, both in the full ABD sample and additional replication cohorts.

INTRODUCTION

Recent years have seen a dramatic increase in the prevalence of obesity in developed countries (32), with reports from the National Center for Health Statistics indicating over 35% of American adults and 17% of children and adolescents are obese (2). Childhood obesity is a serious public health problem that is associated with both immediate and long-term health consequences including increases in blood pressure, cholesterol and insulin resistance as well as social and psychological problems (22-25). Furthermore, research has demonstrated that obese children are more likely to become obese adults (26-30), which is associated with considerable morbidity and mortality including many leading causes of death in developed nations such as diabetes, heart disease and some types of cancer (2, 30, 225). Adolescence represents an important developmental period in which to study obesity because during this time there are rapid changes in physical growth, maturation, and nutritional needs as well as many health-related behaviors are established. Further research is warranted to understand the dynamic process of genetic and environmental influences on BMI from adolescence into adulthood.

Twin and family studies have shown consistently that relative body weight is under considerable genetic influence both in children and adults, with heritability estimates ranging from 50% to 90% (35, 51-54). There have been numerous twin studies examining genetic and environmental influences on adolescent BMI and obesity (35, 54, 55, 57-60, 226-241). However, only five of these studies have reported across the entire time-span of adolescence into adulthood (54, 55, 231, 238, 240). Two large twin study meta-analyses on BMI from birth to young adulthood have reported on over 12,000 twin pairs and found that the contribution of additive genetic effects (A) tend to increase over time while environmental factors common to family members (C) is greatest in childhood but diminishes in adolescence between the ages of 13 and 17 (54, 55). While impressive on scale, these studies do not address the architecture of these effects (i.e., number of factors, persistence across time). Three other studies that reported across adolescence, while longitudinal in design, applied only limited models (Cholesky parameterization), which do not quantify the relative proportion of factors that are common across time versus those that are time specific (242, 243) or examine variance components on rate of BMI change over time (244, 245). Therefore, further twin studies examining alternative models of the genetic and environmental structure across adolescence and into adulthood are warranted.

Genome-wide association studies (GWAS) of BMI using large-scale adult samples have yielded 32 robustly associated genetic variants (63-65) accounting for 1.45% of the phenotypic variation in BMI (63). In a meta-analysis by Speliotes *et al.*, the adult BMI-associated variants were also tested for association in sub-samples of children and adolescents. Based on case/control studies of extreme childhood obesity (n = 1,301-12,891), the authors found nine variants associated with obesity (after correction for multiple testing), including single nucleotide polymorphisms (SNP) in and near *FTO*, *TMEM18* and *MC4R;* in population based samples (n = 354-8,540), three obesity-associated variants were identified in or near *POMC*, *CADM2* and *TNNI3K*; and in parent-child trios with one extreme obese child, the transmission disequilibrium test (TDT) indicated that only alleles in FTO were significantly over-transmitted to obese

children, however, 24 of the 32 effect sizes were in the expected direction (63). Furthermore, a study of 1,097 extreme obese and 2,760 lean controls aged 2-18, found 9 of the 32 adult BMI variants associated with increased risk of obesity including variants in and near *FTO, TMEM18, NRXN3, MC4R, SEC16B, GNPDA2, TNNI3K, QPCTL*, and *BDNF* and also reported 28 variants that were directionally consistent (246). Although, somewhat underpowered, these results indicate adult BMI-associated variants may also be important in childhood and adolescent obesity. A recent GWAS meta-analysis of 5,530 obese and 8,318 control children and adolescents aged 2-18, Bradfield *et al*. reported nine variants significantly associated with obesity. Of these, 7 were previously shown to be associated with adult BMI (*FTO, TMEM18, POMC, MC4R, FAIM2, TNNI3K* and *SEC16B*) and two were in novel loci for childhood obesity (*OLFM4* and *HOXB5*) (247).

While the aforementioned studies implicate a number of genetic variants associated in childhood, adulthood and potentially across the lifespan, they do not address when in development genetic effects begin to influence relative body weight. Therefore, we sought to utilize a developmental twin study design in order to determine the genetic and environmental architecture of BMI by variance components analysis and assess the effects of adult-validated BMI-SNPs across adolescence into adulthood. BMI was calculated from weight and height collected on up to three waves of data collection and ages ranging from 8 to 18 in 2,794 twin participants from the Virginia Twin Study of Adolescent Behavioral Development (ABD).

METHODS

Participants

Participants were from the Virginia Twin Study of Adolescent Behavioral Development (ABD), a longitudinal population-based twin study of adolescent psychopathology. Ascertainment and data collection have been described previously in detail (248-250). In brief, Caucasian twins aged 8 to 17 were recruited through Virginia schools and were followed-up every 18 months for up to three waves of data collection. Of 1,894 eligible Virginia families, 1412 participated in the first wave of data collection (74.5%); 1,047 of 1,302 families that continued to meet the age and Virginia residence requirements completed a second home interview (80%); 628 of 777 eligible families (81%) participated in a third wave of assessment. BMI was calculated from weight and height measurements were collected by trained field interviewers during home interviews who followed a standard protocol using portable scales and tape measures and was available for 2,794 of the ABD twin participants (54% female). For sufficient number of observations over time, age was binned to create five time points: 8-10, 11-12, 13-14, 15-16 and 17-18. If BMI data was collected more than once within a time interval then the average of the assessments was used.

Genotyping

In total, there were 913 participants from 639 families (291 twin pairs, 348 singletons) genotyped on the Illumina Human 660 array. Our quality control procedures removed 2619 monomorphic SNPs, 19984 markers with minor allele frequency less than 1%, 23114 SNPs with greater than 1% missing data and 14 SNPs which deviated from Hardy-Weinberg Equilibrium ($p < 10^{-6}$). Following these exclusions 497,153 genotyped markers remained for analysis. To reduce the effects of population stratification, principal components (PC) were constructed using EIGENSOFT 3.0 (192) and SMARTPCA (193). Because the ABD sample includes related individuals, standard PC analysis s subject to bias. Therefore, the HapMap3 reference panel (988 individuals from 11 human populations) (251) was used to determine SNP weights for each eigenvector and the ABD data was projected onto these values to generate PCs. As recommended by *Patterson et al*., SNPs used to construct PC scores were pruned at $r^2 > 0.7$ to correct for dependence between markers, thereby avoiding disruption of the eigenvalue structure (193). A total of 254,680 autosomal SNPs were used to generate 10 PCs. To circumvent over-fitting, only the first two PCs, distinguishing European from African ancestry, were used in subsequent analyses (192-194).

Genetic risk sum score

BMI SNPs were catalogued from a large-scale BMI meta-analyses by *Speliotes et al.* (63), with 32 SNPs identified as reaching genome-wide significance ($p<5x10^{-8}$). Of the 32 validated BMI SNPs, 15 did not appear on the ABD Illumina 660 array. Therefore, highly-correlated SNPs ($r^2>0.7$) that appeared on the array were used as proxies for ungenotyped SNPs. Proxies for the missing SNPs were identified using SNP Annotation and Proxy Search V2.1 (204), except rs11847697 and rs13107325, for which proxies were unavailable. Haploview v4.10 was used to verify phase and corresponding proxy alleles (148, 149). BMI-associated SNPs were tested in aggregate by constructing GRSSs. There are primarily two approaches for constructing genetic scores: count and weighted methods. The count method is the summation of the number of risk alleles, whereas the weighted method incorporates the sum of the number of risk alleles each weighted by its odds-ratio or effect size. This study utilized the weighted method and constructed GRSS from regression coefficients reported by *Speliotes et. al* (63). GRSSs were calculated using the profile option in PLINK (150).

Variance components modeling

The use of family data allows the particular sources of trait variance to be estimated. In the classical twin design, covariances of MZ and DZ twins are used to estimate the magnitude of genetic and environmental causes of family resemblance (252). This methodology is premised upon monozygotic, or "identical", twins (MZ) sharing all of their genes, while dizygotic, or "fraternal", twins (DZ) sharing half of their genes on average, and MZ and DZ twins sharing environmental experiences to the same extent (equal environment assumption). Following this logic, the correlation between genetic components is modeled as1.0 for MZ twins and 0.5 for DZ twins. Under the assumptions of random mating, no genotype-environment correlation or interaction, and equal environments for MZ and DZ twins, a greater similarity between MZ versus DZ twins is

attributed to additive genetic effects (A). Common environmental effects, as defined in biometrical twin modeling, refer to environmental influences that make family members more similar to each other. Therefore, by definition, these influences correlate 1.0 between both MZ and DZ twins. These shared environmental influences (C) will contribute to twin similarity in both MZ and DZ twins and will tend to increase DZ correlations relative to MZ correlations. However, non-additive genetic effects, known as dominance (D), tend to reduce the DZ correlation relative to MZ twins. The correlation of D is modeled as 1.0 between MZ twins and 0.25 for DZ twins. An additional source of variance is the unique environment (E), which includes factors in the environment that are not shared within families as well as random measurement error. Unique environmental influences are uncorrelated between co-twins and have the effect of decreasing the covariance between siblings. Furthermore, the principles of variance decomposition for the univariate case may be extended to estimating the covariance structure between multiple variables.

One approach to partitioning variance is to use structural equation modeling (SEM) (system of linear equations) and path analysis, which allows for flexible specification of models that include both latent (unobserved) and measured variables (253). In this study, we used SEM to examine the genetic and environmental architecture of BMI across adolescence development. As depicted in Figure 8, independent pathway (IP) models were specified to partition phenotypic variance into genetic and environmental factors that were shared across development as well as components that were time specific (243, 253). These models allow for the contributions of the common factors on the phenotypes measured over time to be different for each of the sources of variance, hence the name 'independent pathways'. ACE models, as opposed to ADE models, were fit as previous research has found shared environment to be important in adolescent BMI (35, 53, 54, 231-236, 238, 241, 254, 255) and upon inspection of the ABD data, the DZ correlations tended to be greater than half MZ correlations which is suggestive of common environmental effects. IP model fitting began with two common factors for each source of variance, A, C and E, along with specific A, C and E at each time point. To simplify the full model, A and C common and specific factors and E common factors were dropped one-by-one from the model. Specific unique environmental effects were not dropped as these include errors of measurement. Variance components models were fit separately by sex and parameters were estimated by full information maximum likelihood using OpenMx (256) in R (210). The log likelihood (-2LL) and Akaike's Information Criterion (AIC) were used to assess goodness-of-fit and relative parsimony of alternative models.

The collective effect of adult BMI-associated genetic variants on BMI across adolescent development was tested via a GRSS (see METHODS *Genetic risk sum score*). The GRSS was added to the best fitting ACE-IP model and was tested as an effect on mean BMI at each time point and separately as an effect on each of the latent genetic factors (Figure 12). To reduce the effects of population stratification, PC scores representing ancestral background were included as covariates in the models as an effect on the mean. As these models include covariates as definition variables, only twins with non-missing values may be used in the analyses. As a consequence, the effective sample size was reduced considerably (Table 18). However, including phenotypic data on ungenotyped relatives has been shown to improve statistical power to detect effects of

genetic variants, as a finite mixture distribution may be used to estimate the probability of genotypes of those ungenotyped (257, 258). Although applying the mixture distribution approach represents the ideal method for this data, we tested the effect of the GRSS by two alternative methods in order to generate preliminary results. First, analyses were performed on an unrelated subset of the ABD twin sample for whom genotyping data was available. Path estimates of the best fitting IP model from the full twin sample were entered as fixed effects while the means, the regression on the PCs and the effects of the GRSS were estimated on mean BMI at each age and separately as an effect on each of the latent genetic factors. Second, we used the parameterization of the best-fit IP model from the full sample but allowed the ACE variance components to be estimated as well as the effect of the GRSS on a sub-sample of ABD twin pairs that were both genotyped. For each method, ten models were evaluated for each sex. Model I estimated the means at each age for the specified model (best-fit IP model) and Model II included the effect of PC covariates on Model I and was considered the baseline for subsequent model comparisons. Model III-VII added the effect of the GRSS separately at each age while Model VIII included the effect at all ages. Model IX included the effect of the GRSS on the first latent common genetic factor and Model X included the effect on the second common genetic factor. The significance of the score was evaluated by comparing models that included the effect of the GRSS and those without and goodness-of-fit of alternative models were assessed by -2LL and AIC.

RESULTS

Descriptive statistics

Means and variances of BMI across age groups are presented by sex and zygosity in Table 14. Females tended to have greater BMI than males at younger ages, while mean BMI for males and females were similar in older age groups (Figure 6). As depicted in Figure 7, the phenotypic variance of BMI tended to increase over time in both males and females, with the largest variance at age 17 for females.

Twin model fitting

The full IP model included two common factors for A, C and E components as well as specific A, C and E components for each of five time points across adolescent development (age 8-18). Model fit and parameter estimates for full and reduced models appear in Table 15 and Table 16. In both females and males, the second C factor and all C specifics could be dropped without significant loss in model fit (Model II.c). According to AIC, the best fitting parameterization of the common C factor featured loadings on age groups 11-16 in females and 8-14 in males (Model II.e). Except for the loadings on age 11 in females, none of the common A factor loadings could be dropped (Model III). However, some of the specific A components could be dropped including age groups 8 and 17 in females and males and, additionally, age group 13 in males (Model V). Furthermore, the second E factor could be dropped in both sexes (Model IV) without significant loss in model fit. Partial path diagrams for best-fit models are displayed in

Figure 9 and Figure 10. BMI was found to be highly heritable, accounting for 74-91% of the variance over the course of adolescent development. The total heritability and proportion of heritability due to common and specific genetic factors for BMI across adolescence are displayed in Table 17 and Figure 11. The proportion of phenotypic variance accounted for by common and specific ACE factors are displayed by age in Figure 11. In summary, the first common genetic factor, which loaded on all time points, tended to account for less of the variance over time from 88 to 41% in females and 74 to 39% in males while the second common genetic factor tended to increase over time from 15 to 49% between ages 13-17 in females and 8 to 50% between ages 11-17 in males. At age 11, 18% and 23% of the heritability was due to a specific genetic factor in females and males, respectively; 14% and 0% at age group 13; and 1% and 8% at age group 15. Thus, the majority of the genetic variance is accounted for by factors that contribute across the adolescent timeframe. Additionally, a common C factor accounted for 2-18% of the phenotypic variation in females from age 11 to 16 and 1-6% in males from age 8 to 14. Furthermore, a common E factor was significant across development and accounted for 2-6% of the phenotypic variance in females and 5-13% in males and specific E factors at each time point accounted for 2-10% of the variance.

Genetic risk sum score (GRSS)

To understand the importance of adult BMI-associated genetic variants across adolescent development, variants were tested collectively by using a GRSS with an effect on each of the common genetic factors and on mean BMI at each age by two alternative methods. The first method assessed the effect of the GRSS in a subsample of unrelated ABD twins (359 females, 258 males) against the background of fixed genetic and environmental factors estimated from the full twin sample (2,794 twins, 54% female). The best fitting model according to goodness-of-fit statistics for both females and males, was Model VIII, which included the effect of the GRSS at each age. Results of model fitting appear in Table 18. The regression coefficients for the GRSS at each age ranged in effect from 0.05 to 1.7 kg/m^2 change in BMI and were in the expected direction (positive, BMI increasing). Next, we assessed the effect of the GRSS while simultaneously estimating genetic and environmental factors in a subsample of genotyped ABD twin pairs (242 female pairs, 152 male pairs). In agreement with the first method, the best fitting model, according to goodness-of-fit statistics, was Model VIII, which included the effect of the GRSS at each age. The results of model fitting are reported in Table 19. The regression coefficients for the GRSS at each age are in the expected direction and ranged in effect from 0.5 to 2.4 kg/m^2 change in BMI. However, the best fitting model according to the AIC, which accounts for model parsimony, differed for males and females; for females, Model VI was the best-fitting model, which only included the effect of the GRSS at age 15 (-0.55 change in BMI); for males, the best-fitting model was IX, which included the effect of the GRSS on the first genetic factor (0.74 kg/m^2 change in BMI). Linear regression indicated that the GRSS accounted for 1-2.3% of the phenotypic variance on BMI across adolescence.

DISCUSSION

The purpose of this study was to utilize a developmental twin study design in order to determine the genetic and environmental architecture of BMI by variance components analysis and to assess the effects of adult-validated BMI SNPs across adolescence. Consistent with other twin and family studies (35, 54, 55, 57-60, 226-241), BMI was found to be highly heritable in the ABD sample, accounting for 74-91% of the variance over the course of adolescent development.

To date, only limited models of the genetic and environmental architecture of BMI have been applied across adolescent development (35, 36, 53). To extend results reported in the literature, independent pathway models were fit to examine genetic and environmental factors, which persisted across time, as well as, time specific. The best-fitting model indicated multiple genetic factors that contributed to BMI liability, including a factor that loaded across development, a second common genetic factor that loaded later in adolescence, and time-specific genetic factors important during mid-adolescence (ages 11 to 15). It is possible that these specific genetic components are reflective of genetic effects related to puberty. Puberty stage has been shown to be highly heritable (259) and to have a significant effect on BMI variance, with higher genetic variance at later pubertal stages (240). The findings reported here do not incorporate effects of puberty and are likely confounded by the use of chronological age without consideration of puberty stage. Accordingly, our forthcoming analyses will incorporate the effects of puberty on adolescent BMI development.

Our results indicated that shared environmental effects accounted for a portion of the phenotypic variance in adolescent BMI (1-18%), although timing differed between the sexes, with significant effects until ages 14 and 16 in males and females, respectively. These results were consistent with other twin studies which report environmental effects shared within families to be important for BMI, as well as, confirming these effects diminished in adolescence between the ages of 13 and 17 (54, 55). Additionally, our results indicated a common unique environmental factor, which loaded across development, accounting for 2-13% of the phenotypic variance in BMI. These results suggested that there were environmental factors specific to individuals that persisted across time to influence body composition. These results further supported the importance of environmental factors, both within families and specific to individuals, contributing to the progression of relative body weight. Previous research has identified specific environmental factors shown to influence obesity including food selection, physical activity, socioeconomic status and childhood abuse (260-268). For example, the heritability of BMI has been shown to decrease with high physical activity (260-262). Future research should incorporate known environmental moderators into variance decomposition modeling to further clarify the genetic and environmental architecture and tracking of relative body weight across the lifespan.

In addition, to investigate the effect of adult BMI-associated genetic variants in adolescence, variants were tested as a collection by a GRSS with an effect on each common genetic factor and on mean BMI at each age by two alternative methods. To the best of our knowledge, this is the first study to examine the association of adult BMI-variants across adolescence assessed within the context of genetic and environmental components determined by variance decomposition. Preliminary results, evaluated using

subsamples of ABD twins, indicated that the GRSS was best modeled as an effect on mean BMI at each age group, suggesting association across development, with the magnitude of the effect differing at each time point considered and ranged in effect from 0.05 to 2.4 kg/m^2 change in BMI.

The initial GRSS results reported here should be interpreted in light of several limitations. First, since only a portion of the ABD sample was genotyped, association analyses were performed on a reduced sample of twins, limiting our power to detect significant associations. Despite reported test-statistics reflecting improvement in model-fit with the addition of genetic scores, confidence intervals on the corresponding effect-sizes were large and often inclusive of zero, indicating the need for larger sample sizes to resolve the nature of these effects. The inclusion of DZ twin pairs of opposite sex (DZo) in subsequent analyses would increase the effective sample size, as well as, allow for statistical examination of sex effects. Indeed, genetic epidemiology studies of adolescent body composition support the presence of sex limitation (51, 54, 56-60), as do molecular genetic studies (269, 270). Additionally, research indicates that including phenotypic data on ungenotyped relatives improves statistical power to detect effects of genetic variants, as a finite mixture distribution may be used to estimate the probability of unobserved genotypes in untyped individuals (257, 258). Thus, extensions to this work will not only include DZo twins to track sex effects in BMI across adolescence, but also incorporate mixture distribution methodology, to increase power to potentially detect relevant associations.

Additionally, our results found that a GRSS comprised of 30 adult BMI-associated genetic variants accounted for 1-2.3% of the phenotypic variance in BMI across adolescence. Other studies examining genetic risk scores in children incorporated 8 to 17 risk variants and found them to account for 0.8 to 2.2% of the phenotypic variance in BMI (271-275). To date, no studies of adolescent body weight have incorporated genetic risk scores in the context of twin methodology and variance decomposition. However, one longitudinal twin study by Haworth *et al.*, modeled the genetic and environmental architecture of BMI in children aged 4 to 11 by Cholesky decomposition and then separately examined the effect of a variant in *FTO* in a subset of unrelated twins. The authors reported that the SNP accounted for 0.1% of the variance at age 4 and increased over time to 1.0% by age 11 (234). There is a need for additional research examining the effects of validated obesity loci across development.

A number of extensions may be applied to the genetic sum score methodology presented herein. For example, other classes of genetic variation such as copy number variation (CNV), insertions, deletions and lower-frequency SNPs may be incorporated into genetic profiles, as well as comparison of methods based on allelic count versus weights. For example, our group has previously examined CNVs reported to be associated with BMI and obesity (Chapter 3), replicating an association with a deletion on 16p12.3 in an adult sample. Future research should examine these variants in samples of children and incorporate these into genetic burden scores. In addition, there are various other latent variable models that may be applied in conjunction with genetic risk scores to expand insight on the development of relative body weight. Potential models include simplex and growth curves, which would allow the assessment of variance components and genetic variants on innovations, transmissions and rate of change of BMI across time.

In summary, we have utilized a developmental twin study design to examine the genetic and environmental architecture of BMI by variance components analysis. We found BMI to be highly heritable accounting for 74-94% of the variance across adolescence, which was reflected by several genetic factors associated across time and at specific ages, as well as environmental factors, both common to family members and specific to individuals, persisting across development. Furthermore, we assessed the effects of adult-validated BMI-SNPs across adolescence within the context of genetic and environmental factors determined by variance decomposition. Our results indicated that the GRSS was associated across development and accounted for 1-2.3% of the phenotypic variance in BMI across adolescence. These findings, although preliminary, merit future research, which considers pubertal stage, both in the full ABD sample and additional replication cohorts. Understanding obesity development will aid in identifying obesogenic vulnerability time-points and facilitate targeted prevention and treatment efforts.

TABLES AND FIGURES

Table 14: Descriptive statistics for BMI by zygosity and age group

	FEMALES				MALES			
Age Zyg	**Mean T1 (Var)**	**Mean T2 (Var)**	**Cor (Cov)**	**Pairs/ Singles**	**Mean T1 (Var)**	**Mean T2 (Var)**	**Cor (Cov)**	**Pairs/ Singles**
8 MZ	17.80 (8.38)	17.64 (8.51)	0.86 (7.29)	141/3	17.13 (8.30)	16.94 (7.29)	0.74 (5.74)	87/1
8 DZ	18.57 (17.50)	17.47 (11.53)	0.52 (7.38)	58/1	17.28 (11.35)	17.49 (13.61)	0.45 (5.58)	64/2
11 MZ	18.97 (12.14)	18.87 (11.19)	0.87 (10.13)	159/0	19.05 (13.18)	18.50 (12.03)	0.92 (11.53)	132/1
11 DZ	20.04 (19.76)	19.55 (22.01)	0.57 (11.80)	82/1	18.89 (18.71)	18.30 (17.18)	0.59 (10.49)	81/0
13 MZ	20.20 (12.79)	20.25 (14.07)	0.88 (11.80)	199/1	20.34 (13.83)	19.93 (13.63)	0.90 (12.37)	165/2
13 DZ	22.33 (25.50)	21.64 (22.91)	0.58 (14.04)	101/2	20.63 (21.97)	19.80 (26.84)	0.59 (14.26)	92/0
15 MZ	21.60 (17.87)	21.42 (14.96)	0.88 (14.45)	223/3	21.30 (11.99)	21.03 (10.86)	0.87 (9.92)	170/2
15 DZ	23.29 (23.96)	22.31 (19.39)	0.58 (12.55)	80/0	22.22 (17.02)	21.97 (23.05)	0.44 (8.78)	98/2
17 MZ	22.14 (27.86)	21.95 (22.15)	0.9 (22.39)	124/4	23.09 (18.83)	22.89 (21.84)	0.9 (18.32)	120/1
17 DZ	23.40 (28.03)	22.59 (16.54)	0.58 (12.55)	49/3	23.01 (22.10)	22.87 (14.39)	0.41 (7.35)	50/4

Note: BMI = body mass index, Age = age group in years, Zyg = zygosity, MZ = monozygotic, DZ = dizygotic, T1 = twin one, T2 = twin two, Var = variance, Cov = covariance, Cor = within-pair Pearson correlation coefficient, Pairs = number of complete twin pairs, Singles = number of twin singletons.

Figure 6: Mean BMI by sex and age group in the ABD sample

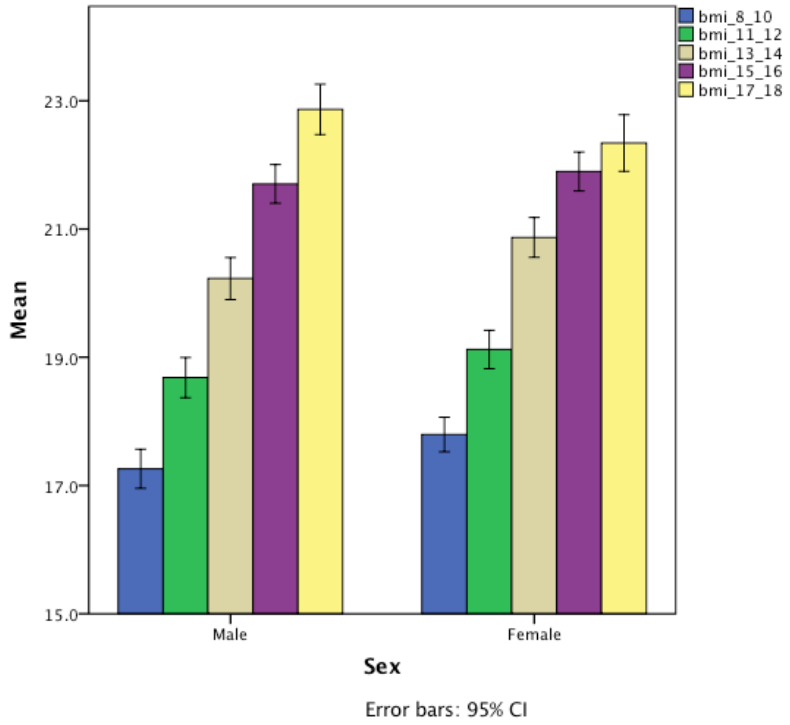


Figure 7: Variance BMI by sex and age group in the ABD sample

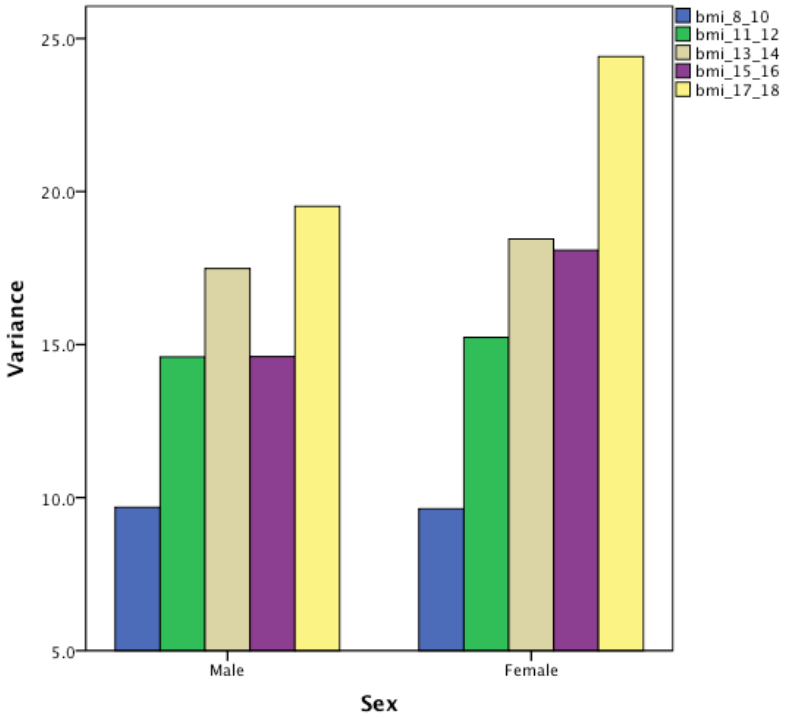


Table 15: IP-1f Females and Males ABD

FEMALES: IP_1f

Model	EP	-2LL	df	AIC	Diff LL	Diff df	p-value	ac11 stdPathEst stdPath^2 [95%CI]	ac21 stdPathEst stdPath^2 [95%CI]	ac31 stdPathEst stdPath^2 [95%CI]	ac41 stdPathEst stdPath^2 [95%CI]	ac51 stdPathEst stdPath^2 [95%CI]	cc11 stdPathEst stdPath^2 [95%CI]	cc21 stdPathEst stdPath^2 [95%CI]	cc31 stdPathEst stdPath^2 [95%CI]	cc41 stdPathEst stdPath^2 [95%CI]	cc51 stdPathEst stdPath^2 [95%CI]	ce11 stdPathEst stdPath^2 [95%CI]	ce21 stdPathEst stdPath^2 [95%CI]	ce31 stdPathEst stdPath^2 [95%CI]	ce41 stdPathEst stdPath^2 [95%CI]	ce51 stdPathEst stdPath^2 [95%CI]	as11 stdPathEst stdPath^2 [95%CI]	as22 stdPathEst stdPath^2 [95%CI]	as33 stdPathEst stdPath^2 [95%CI]	as44 stdPathEst stdPath^2 [95%CI]	as55 stdPathEst stdPath^2 [95%CI]	cs11 stdPathEst stdPath^2 [95%CI]	cs22 stdPathEst stdPath^2 [95%CI]	cs33 stdPathEst stdPath^2 [95%CI]	cs44 stdPathEst stdPath^2 [95%CI]	cs55 stdPathEst stdPath^2 [95%CI]	es11 stdPathEst stdPath^2 [95%CI]	es22 stdPathEst stdPath^2 [95%CI]	es33 stdPathEst stdPath^2 [95%CI]	es44 stdPathEst stdPath^2 [95%CI]	es55 stdPathEst stdPath^2 [95%CI]
I. IP-1f	35	11385.07	2415	6555.07	-	-	-	2.849	3.265	3.673	4.008	4.168	0.754	0.447	-1.24	-1.494	-2.199	0.422	0.701	1.052	1.043	1.29	0.428	1.866	-1.514	0.885	1.296	0	0	0.963	0	0	-1.031	-1.003	-0.77	-0.962	0.908
								0.896	0.82	0.822	0.873	0.812	-0.237	-0.112	0.278	0.325	0.428	0.133	0.176	0.236	0.227	0.251	0.135	0.469	0.339	0.193	0.252	0	0	0.216	0	0	-0.324	-0.252	-0.172	-0.209	0.177
								0.803	0.673	0.676	0.762	0.659	0.056	0.013	0.077	0.106	0.183	0.018	0.031	0.056	0.052	0.063	0.018	0.22	0.115	0.037	0.064	0	0	0.047	0	0	0.105	0.063	0.03	0.044	0.031
								[0.588,0.895]	[0.532,0.756]	[0.52,0.777]	[0.57,0.899]	[0.454,0.831]	[0,0.228]	[0,0.153]	[0,0.243]	[0.001,0.289]	[0.015,0.399]	[0.003,0.044]	[0.015,0.053]	[0.035,0.082]	[0.032,0.079]	[0.039,0.094]	[0,0.133]	[0.143,0.303]	[0.035,0.202]	[0,0.086]	[0,0.124]	[0,0.042]	[0,0.031]	[0,0.142]	[0,0.026]	[0,0.08]	[0.075,0.144]	[0.044,0.091]	[0.014,0.046]	[0.026,0.063]	[0.015,0.05 1]
II. Drop Specific C	30	11385.77	2420	6545.77	0.71	5	0.98	2.883	3.276	3.63	3.941	4.098	-0.682	-0.333	1.425	1.653	2.34	-0.421	-0.7	-1.055	-1.047	-1.29	0.293	1.868	1.746	0.925	1.308	-	-	-	-	-	1.03	-1.001	-0.762	-0.96	-0.908
								0.907	0.824	0.813	0.857	0.797	-0.215	-0.084	0.319	0.359	0.455	-0.132	-0.176	-0.236	-0.228	-0.251	0.092	0.47	0.391	0.201	0.254	-	-	-	-	-	0.324	-0.252	-0.171	-0.209	-0.177
								0.823	0.678	0.66	0.735	0.634	0.046	0.007	0.102	0.129	0.207	0.018	0.031	0.056	0.052	0.063	0.009	0.221	0.153	0.04	0.065	-	-	-	-	-	0.105	0.063	0.029	0.044	0.031
								[0.629,0.898]	[0.562,0.758]	[0.508,0.773]	[0.555,0.88]	[0.44,0.807]	[0,0.199]	[0,0.121]	[0.003,0.258]	[0.008,0.303]	[0.032,0.413]	[0.004,0.044]	[0.015,0.053]	[0.036,0.082]	[0.032,0.079]	[0.039,0.094]	[0,0.123]	[0.143,0.303]	[0.106,0.208]	[0,0.088]	[0,0.124]	-	-	-	-	-	[0.075,0.142]	[0.044,0.09]	[0.013,0.045]	[0.026,0.063]	[0.015,0.05 1]
III. Drop as11,as44	28	11388.89	2422	6544.89	3.82	7	0.8	2.86	3.248	3.658	4.068	4.102	-0.804	-0.452	1.24	1.616	2.204	-0.417	-0.711	-1.071	-1.007	-1.289	-	-1.897	-1.824	-	1.512	-	-	-	-	-	1.049	0.998	0.736	1.027	-0.897
				II.	3.12	2	0.21	0.9	0.816	0.819	0.883	0.798	-0.253	-0.114	0.278	0.351	0.429	-0.131	-0.178	-0.24	-0.219	-0.251	-	-0.476	-0.408	-	0.294	-	-	-	-	-	0.33	0.251	0.165	0.223	-0.174
								0.81	0.665	0.671	0.78	0.636	0.064	0.013	0.077	0.123	0.184	0.017	0.032	0.058	0.048	0.063	-	0.227	0.167	-	0.086	-	-	-	-	-	0.109	0.063	0.027	0.05	0.03
								[0.638,0.891]	[0.549,0.744]	[0.529,0.766]	[0.598,0.899]	[0.443,0.794]	[0.001,0.227]	[0,0.121]	[0.002,0.22]	[0.009,0.301]	[0.022,0.393]	[0.004,0.043]	[0.015,0.054]	[0.037,0.085]	[0.029,0.074]	[0.039,0.094]	-	[0.157,0.309]	[0.125,0.218]	-	[0.023,0.139]	-	-	-	-	-	[0.081,0.146]	[0.043,0.09]	[0.01,0.044]	[0.033,0.069]	[0.013,0.05 1]
IV. Drop cc21	27	11389.8	2423	6543.8	4.73	8	0.79	2.93	3.271	3.528	3.894	3.902	-0.481	-	1.647	2.053	2.579	0.412	0.702	1.073	1.022	1.3	-	1.906	-1.828	-	1.543	-	-	-	-	-	1.038	0.999	-0.738	1.031	0.893
				III.	0.91	1	0.34	0.924	0.822	0.785	0.84	0.755	-0.152	-	0.366	0.443	0.499	0.13	0.176	0.239	0.221	0.251	-	0.479	-0.407	-	0.298	-	-	-	-	-	0.327	0.251	-0.164	0.222	0.173
								0.853	0.676	0.616	0.706	0.569	0.023	-	0.134	0.196	0.249	0.017	0.031	0.057	0.049	0.063	-	0.23	0.165	-	0.089	-	-	-	-	-	0.107	0.063	0.027	0.049	0.03
								[0.798,0.892]	[0.591,0.749]	[0.536,0.688]	[0.608,0.789]	[0.436,0.707]	[0.003,0.064]	-	[0.073,0.208]	[0.118,0.29]	[0.105,0.408]	[0.004,0.042]	[0.015,0.053]	[0.036,0.084]	[0.029,0.076]	[0.039,0.094]	-	[0.158,0.312]	[0.124,0.217]	-	[0.031,0.139]	-	-	-	-	-	[0.08,0.143]	[0.044,0.09]	[0.01,0.043]	[0.032,0.068]	[0.013,0.05 5]

MALES: IP_1f

Model	EP	-2LL	df	AIC	Diff LL	Diff df	p-value	ac11 stdPathEst stdPath^2 [95%CI]	ac21 stdPathEst stdPath^2 [95%CI]	ac31 stdPathEst stdPath^2 [95%CI]	ac41 stdPathEst stdPath^2 [95%CI]	ac51 stdPathEst stdPath^2 [95%CI]	cc11 stdPathEst stdPath^2 [95%CI]	cc21 stdPathEst stdPath^2 [95%CI]	cc31 stdPathEst stdPath^2 [95%CI]	cc41 stdPathEst stdPath^2 [95%CI]	cc51 stdPathEst stdPath^2 [95%CI]	ce11 stdPathEst stdPath^2 [95%CI]	ce21 stdPathEst stdPath^2 [95%CI]	ce31 stdPathEst stdPath^2 [95%CI]	ce41 stdPathEst stdPath^2 [95%CI]	ce51 stdPathEst stdPath^2 [95%CI]	as11 stdPathEst stdPath^2 [95%CI]	as22 stdPathEst stdPath^2 [95%CI]	as33 stdPathEst stdPath^2 [95%CI]	as44 stdPathEst stdPath^2 [95%CI]	as55 stdPathEst stdPath^2 [95%CI]	cs11 stdPathEst stdPath^2 [95%CI]	cs22 stdPathEst stdPath^2 [95%CI]	cs33 stdPathEst stdPath^2 [95%CI]	cs44 stdPathEst stdPath^2 [95%CI]	cs55 stdPathEst stdPath^2 [95%CI]	es11 stdPathEst stdPath^2 [95%CI]	es22 stdPathEst stdPath^2 [95%CI]	es33 stdPathEst stdPath^2 [95%CI]	es44 stdPathEst stdPath^2 [95%CI]	es55 stdPathEst stdPath^2 [95%CI]
I. IP-1fFull	35	9572.71	2098	5376.71	-	-	-	2.28	3.071	3.593	3.85	3.772	1.114	0.733	0.791	0.198	-0.429	1.08	1.067	0.991	0.927	0.978	-0.865	1.626	1.151	0.875	1.574	0	0.55	-1.137	0	0	0.843	-0.523	-0.81	0.87	1.023
								0.757	0.811	0.852	0.927	0.868	0.37	0.194	0.187	0.048	-0.099	0.359	0.282	0.235	0.223	0.225	-0.287	0.43	0.273	0.211	0.362	0	0.145	-0.27	0	0	0.28	-0.138	-0.192	0.209	0.235
								0.574	0.658	0.726	0.86	0.753	0.137	0.038	0.035	0.002	0.01	0.129	0.079	0.055	0.05	0.051	0.083	0.184	0.074	0.044	0.131	0	0.021	0.073	0	0	0.078	0.019	0.037	0.044	0.055
								[0.372,0.775]	[0.421,0.755]	[0.549,0.829]	[0.733,0.924]	[0.567,0.856]	[0,0.352]	[0,0.326]	[0,0.208]	[0,0.115]	[0,0.192]	[0.075,0.199]	[0.051,0.116]	[0.035,0.083]	[0.029,0.077]	[0.024,0.086]	[0,0.2]	[0.098,0.268]	[0.0174]	[0,0.111]	[0.035,0.233]	[0,0.163]	[0,0.115]	[0,0.152]	[0,0.018]	[0,0.042]	[0.044,0.134]	[0.004,0.036]	[0.024,0.052]	[0.029,0.065	[0.035,0.085]
II. Drop C	28	9577.37	2105	5367.37	4.66	7	0.7	2.38	3.124	3.652	3.803	3.688	-0.895	-0.671	-	-	-	1.063	1.061	0.992	0.929	1.01	0.928	-1.65	1.085	-0.984	1.865	-	-	1.239	-	-	-0.827	-0.528	-0.809	-0.862	-0.99
								0.787	0.825	0.868	0.921	0.844	-0.296	-0.177	-	-	-	0.352	0.28	0.236	0.225	0.231	0.307	-0.436	0.258	-0.238	0.427	-	-	0.295	-	-	-0.273	-0.139	-0.192	-0.209	-0.227
								0.62	0.681	0.754	0.849	0.713	0.088	0.033	-	-	-	0.124	0.079	0.056	0.051	0.053	0.094	0.19	0.067	0.057	0.182	-	-	0.087	-	-	0.075	0.019	0.037	0.044	0.051
								[0.517,0.709]	[0.596,0.753]	[0.687,0.84]	[0.781,0.911]	[0.628,0.784]	[0,0.238]	[0,0.132]	-	-	-	[0.074,0.187]	[0.051,0.114]	[0.035,0.083]	[0.03,0.077]	[0.028,0.088]	[0,0.216]	[0.1,0.262]	[0.0163]	[0.001,0.117]	[0.12,0.259]	-	-	[0.006,0.159]	-	-	[0.043,0.126]	[0.005,0.036]	[0.025,0.052]	[0.029,0.064	[0.03,0.078]
III. Drop A	27	9578.29	2106	5366.29	5.58	8	0.69	2.4	3.133	3.658	3.799	3.677	-1.229	-0.457	-	-	-	1.082	1.056	0.994	0.938	0.977	-	1.718	-1.098	-0.994	1.902	-	-	1.221	-	-	-0.864	0.53	-0.806	-0.85	1.01
				II.	0.92	1	0.34	0.792	0.826	0.869	0.921	0.841	-0.405	-0.121	-	-	-	0.357	0.278	0.236	0.227	0.224	-	0.453	-0.261	-0.241	0.435	-	-	0.29	-	-	-0.285	0.14	-0.192	-0.206	0.231
								0.627	0.683	0.755	0.848	0.707	0.164	0.015	-	-	-	0.127	0.078	0.056	0.052	0.05	-	0.205	0.068	0.058	0.189	-	-	0.084	-	-	0.081	0.02	0.037	0.042	0.053
								[0.523,0.715]	[0.601,0.754]	[0.688,0.815]	[0.78,0.909]	[0.621,0.781]	[0.093,0.25]	[0,0.06]	-	-	-	[0.073,0.195]	[0.05,0.113]	[0.035,0.084]	[0.031,0.078]	[0.026,0.083]	-	[0.15,0.269]	[0.0164]	[0.003,0.118]	[0.124,0.267]	-	-	[0.004,0.156]	-	-	[0.048,0.133]	[0.005,0.036]	[0.024,0.052]	[0.028,0.062]	[0.034,0.08]
IV. Drop as33,cc21	25	9583.49	2108	5367.49	10.78	10	0.38	2.466	3.196	3.744	3.691	3.685	-1.103	-	-	-	-	-1.082	-1.072	-0.959	-0.928	-0.993	-	1.689	-	1.211	-1.953	-	-	-1.532	-	-	-0.868	-0.503	-0.854	0.865	-0.996
				III.	5.2	2	0.07	0.812	0.84	0.882	0.903	0.837	-0.363	-	-	-	-	-0.356	-0.282	-0.226	-0.227	-0.226	-	0.444	-	0.296	-0.444	-	-	-0.361	-	-	-0.286	-0.132	-0.201	0.212	-0.226
								0.66	0.706	0.778	0.816	0.701	0.132	-	-	-	-	0.127	0.079	0.051	0.052	0.051	-	0.197	-	0.088	0.197	-	-	0.13	-	-	0.082	0.018	0.041	0.045	0.051
								[0.568,0.737]	[0.633,0.77]	[0.719,0.829]	[0.754,0.871]	[0.613,0.776]	[0.072,0.205]	-	-	-	-	[0.072,0.195]	[0.052,0.1 5]	[0.032,0.078]	[0.03 1,0.079]	[0.026,0.084]	-	[0.143,0.26]	-	[0.039,0.142]	[0.131,0.276]	-	-	[0.087,0.182]	-	-	[0.048,0.133]	[0.003,0.034]	[0.028,0.056]	[0.029,0.065	[0.033,0.077]

Table: IJMALES IP2f

| Model | EP | -2LL | df | AIC | Diff LL | Diff df | p-value | ac11 stdPunExt [95%CI] | ac21 stdPunMxt [95%CI] | ac31 stdPunAst [95%CI] | ac41 stdPunExt [95%CI] | ac51 stdPunMxt [95%CI] | ac22 stdPunExt [95%CI] | ac32 stdPunAst [95%CI] | ac42 stdPunMxt [95%CI] | ac52 stdPunAst [95%CI] | cc11 stdPunAst [95%CI] | cc21 stdPunExt [95%CI] | cc31 stdPunMxt [95%CI] | cc41 stdPunAst [95%CI] | cc51 stdPunMxt [95%CI] | cc22 stdPunExt [95%CI] | cc32 stdPunAst [95%CI] | cc42 stdPunMxt [95%CI] | cc52 stdPunAst [95%CI] | ac11 stdPunExt^2 [95%CI] | ac21 stdPunMxt^2 [95%CI] | ac31 stdPunAst^2 [95%CI] | ac41 stdPunExt^2 [95%CI] | ac51 stdPunMxt^2 [95%CI] | ac22 stdPunExt^2 [95%CI] | ac32 stdPunAst^2 [95%CI] | ac42 stdPunMxt^2 [95%CI] | ac52 stdPunAst^2 [95%CI] | cc11 stdPunAst^2 [95%CI] | cc21 stdPunExt^2 [95%CI] | cc31 stdPunMxt^2 [95%CI] | cc41 stdPunAst^2 [95%CI] | cc51 stdPunMxt^2 [95%CI] | cc22 stdPunExt^2 [95%CI] | cc32 stdPunAst^2 [95%CI] | cc42 stdPunMxt^2 [95%CI] | cc52 stdPunAst^2 [95%CI] | aa11 stdPunExt [95%CI] | aa22 stdPunExt [95%CI] | aa33 stdPunAst [95%CI] | aa44 stdPunMxt [95%CI] | aa55 stdPunAst [95%CI] | cc11 stdPunExt [95%CI] | cc22 stdPunExt [95%CI] | cc33 stdPunAst [95%CI] | cc44 stdPunMxt [95%CI] | aa55 stdPunAst [95%CI] | ee11 stdPunExt [95%CI] | ee22 stdPunExt [95%CI] | ee33 stdPunAst [95%CI] | ee44 stdPunMxt [95%CI] | ee55 stdPunAst [95%CI] |
|---|---|---

Figure 8: Independent pathway diagram for two common ACE factors and specific ACE components for five observed variables

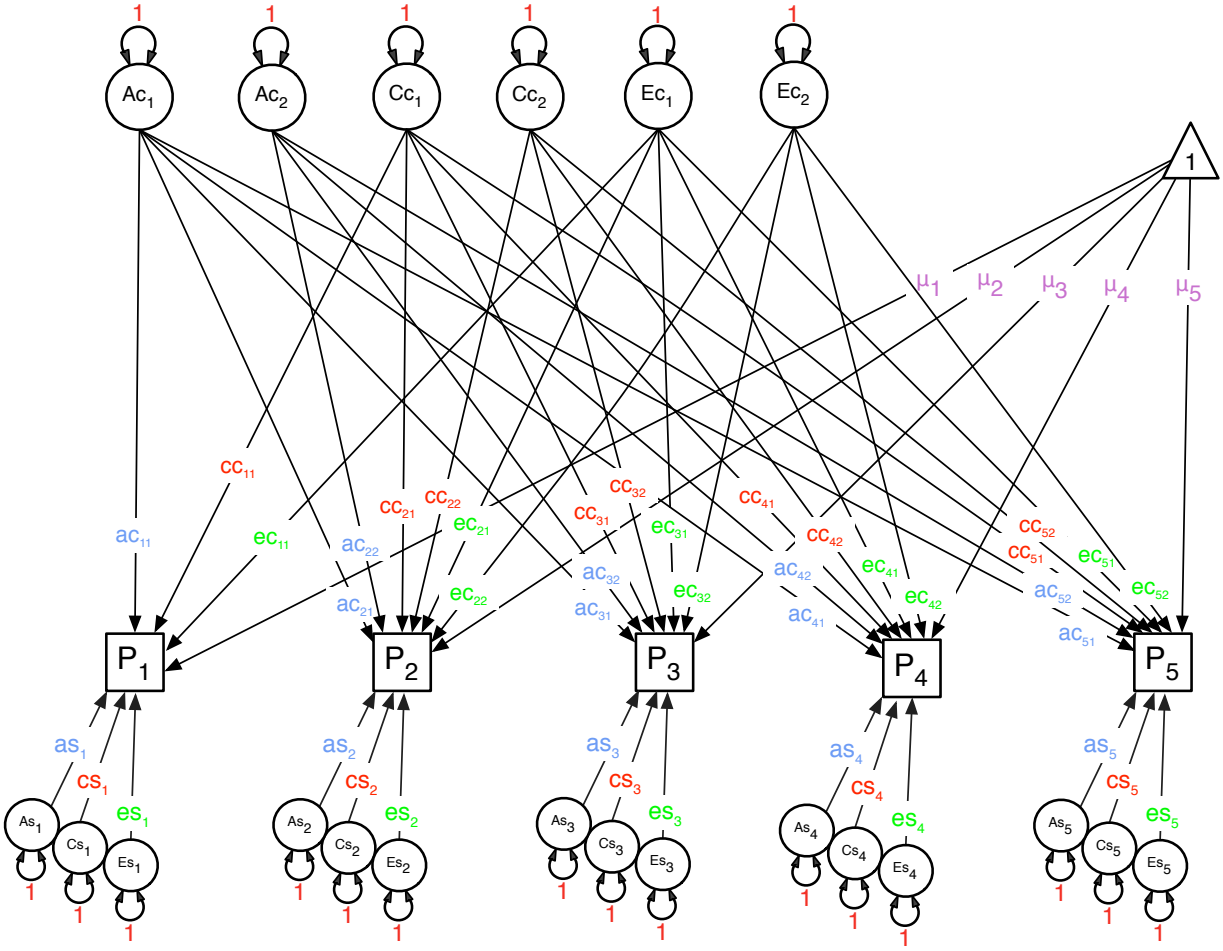


Figure 9: Partial IP path diagram with path estimates for females in the ABD sample

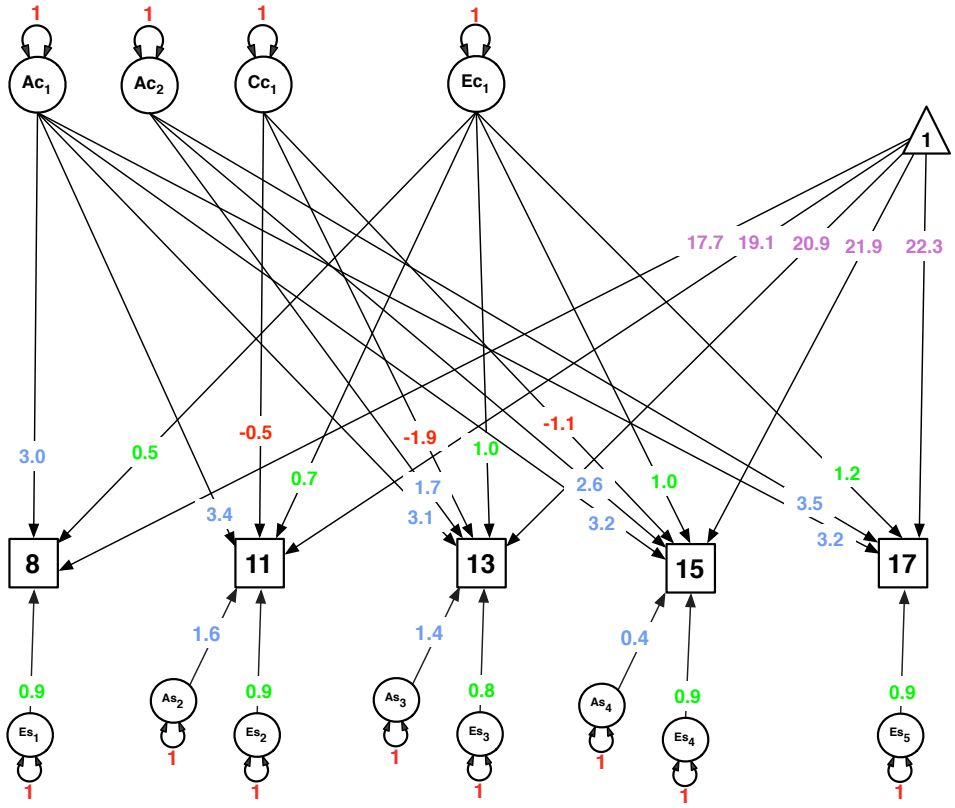


Figure 10: Partial IP path diagram with path estimates for males in the ABD sample

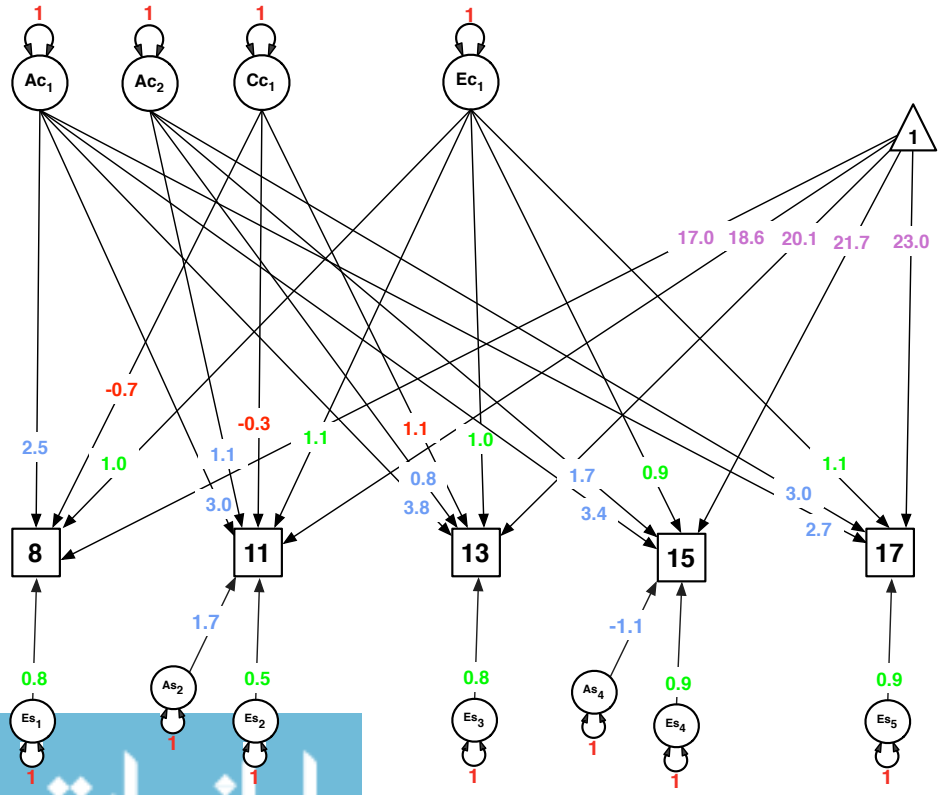
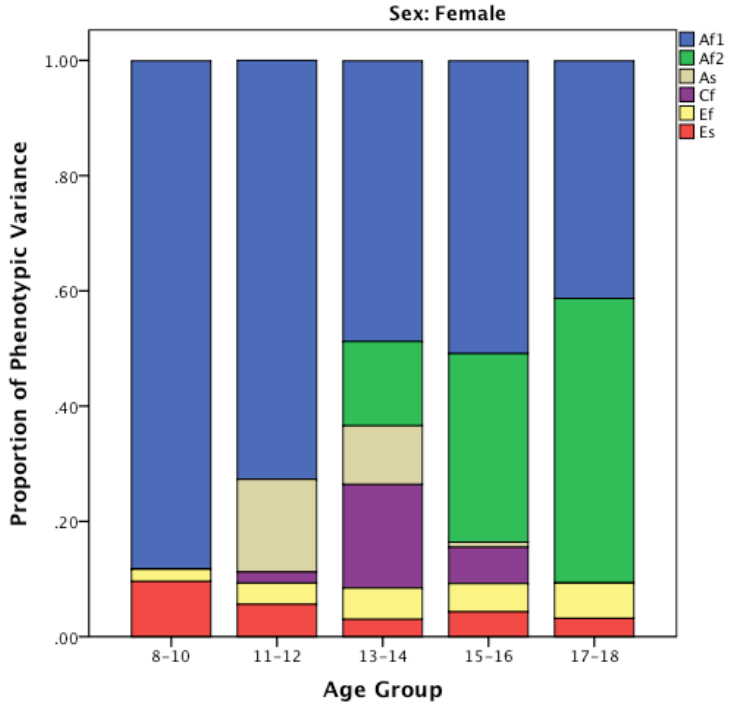


Figure 11: Proportion of phenotypic variance accounted for by common and specific genetic and environmental components by sex

a) Females



b) Males

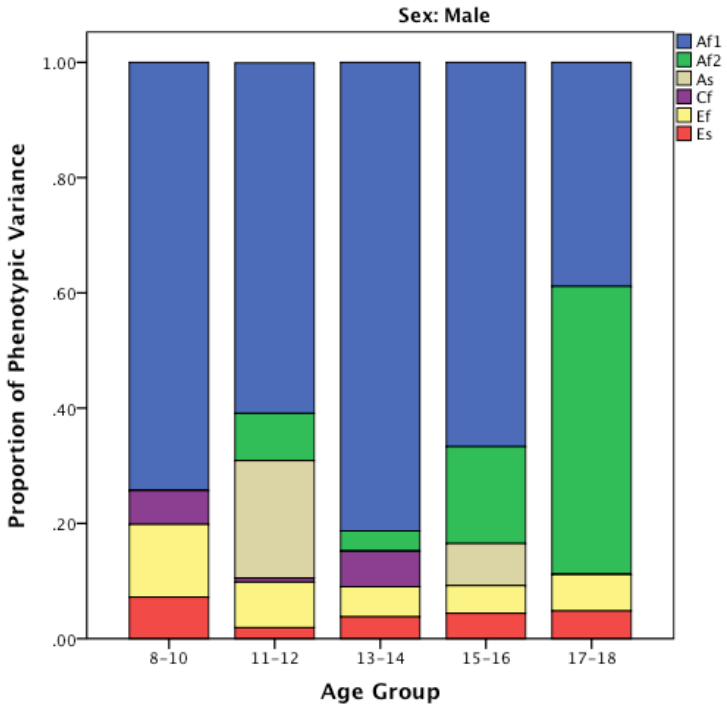


Figure 12: Partial path diagram including effects of GRSS on BMI in females across adolescence

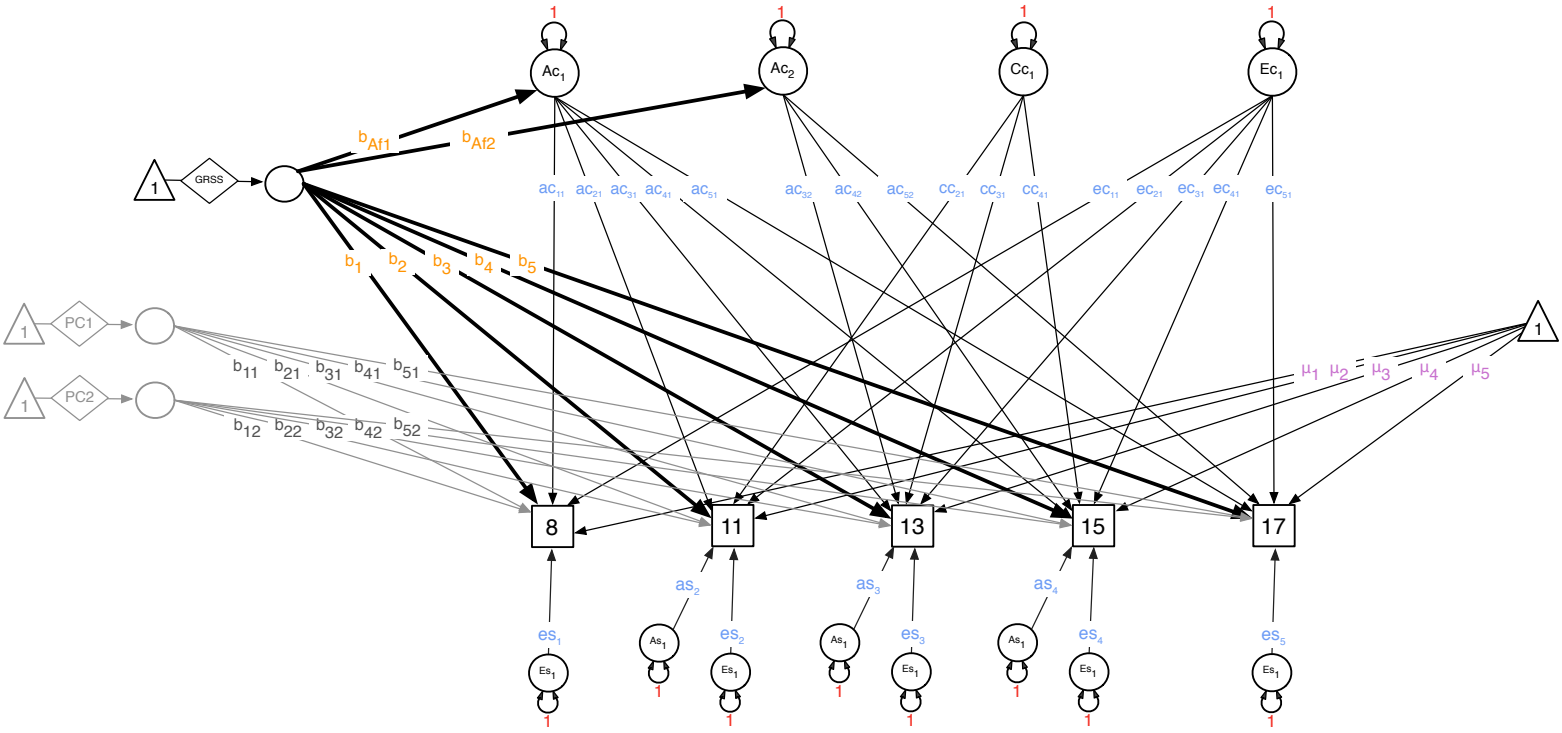


Table 17: Total heritability and proportion of heritability due to common and specific genetic factors for BMI across adolescence

Age Group	Sex	Total Heritability	% A Factor 1	% A Factor 2	% A Specific
8-10	Female	0.88	100	0	0
	Male	0.74	100	0	0
11-12	Female	0.89	82	0	18
	Male	0.89	68	9	23
13-14	Female	0.74	66	20	14
	Male	0.85	96	4	0
15-16	Female	0.84	60	39	1
	Male	0.91	73	19	8
17-18	Female	0.91	45	55	0
	Male	0.89	44	56	0

Note: BMI = body mass index, A = additive genetic component.

Table 18: Effect of GRSS on common genetic factors and mean BMI by sex in an un-related subsample of genotyped ABD participants (Method 1)

Model: Females	EP	-2LL	df	AIC	Diff LL	Difdf	p-value	Beta [95%CI]	Estimated Means
I. Means	5	4471.75	794	2883.75	-	-	-	-	17.9 18.6 20.9 21.9 22.6
II. Means & PCs	15	4412.09	784	2844.09	-	-	-	-	17.8 19.2 20.9 21.9 23.0
III. Age 8-10	16	4410.25	783	2844.25	1.84	1	0.17	0.287 [-0.129,0.704]	18.2 19.2 20.9 21.9 23.0
IV. Age 11-12	16	4411	783	2845	1.1	1	0.3	-0.205 [-0.590,0.180]	17.8 18.9 20.9 21.9 23.0
V. Age 13-14	16	4397.2	783	2831.2	14.89	1	<0.001	0.626 [0.307,0.944]	17.8 19.2 21.9 21.9 23.0
VI. Age 15-16	16	4397.27	783	2831.27	14.83	1	<0.001	-0.488 [-0.736,-0.239]	17.8 19.2 20.9 21.2 23.0
VII. Age 17-18	16	4405.52	783	2839.52	6.57	1	0.01	0.609 [0.142,1.078]	17.8 19.2 20.9 21.9 23.9
VIII. Each age	20	4391.6	779	2833.6	20.5	5	<0.001	0.573 0.743 1.027 0.352 0.651 [-0.157,1.304] [-0.101,1.587] [0.127,1.927] [-0.557,1.261] [-0.439,1.740]	18.6 20.3 22.4 22.4 24.0
VIII.a Age 13-18	18	4394.71	781	2832.71	17.38	3	<0.001	0.424 -0.184 0.155 [-0.151,0.999] [-0.866,0.497] [-0.775,1.085]	17.8 19.2 21.6 21.6 23.2
IX. Common A1	16	4410.15	783	2844.15	1.94	1	0.16	0.159 [-0.065,0.384]	18.5 19.9 21.6 22.6 23.7
X. Common A2	16	4412.09	783	2846.09	<0.001	1	0.98	0.004 [-0.245,0.252]	17.8 19.2 20.9 21.9 23.0

Model: Males	EP	-2LL	df	AIC	Diff LL	Difdf	p-value	Beta [95%CI]	Estimated Means
I. Means	5	2914.75	571	1772.75	-	-	-	-	17.0 18.9 20.7 22.0 22.6
II. Means & PCs	15	2902.49	561	1780.49	-	-	-	-	16.9 18.9 20.3 22.0 22.7
III. Age 8-10	16	2899.06	560	1779.06	3.43	1	0.06	-0.627 [-1.296,0.041]	16.2 18.9 20.3 21.9 22.7
IV. Age 11-12	16	2894.64	560	1774.64	7.85	1	0.01	1.052 [0.311,1.793]	17.0 20.3 20.3 21.9 22.7
V. Age 13-14	16	2902.42	560	1782.42	0.07	1	0.79	0.079 [-0.555,0.714]	16.9 18.9 20.4 21.9 22.7
VI. Age 15-16	16	2894.71	560	1774.71	7.77	1	0.01	0.845 [0.248,1.443]	17.0 18.9 20.3 23.1 22.7
VII. Age 17-18	16	2891.76	560	1771.76	10.73	1	p <0.001	-1.575 [-2.525,-0.625]	17.0 18.9 20.3 22.0 20.4
VIII. Each age	20	2875.93	556	1763.93	26.56	5	p <0.001	0.535 1.763 1.579 1.679 0.046 [-0.372,1.443] [0.700,2.827] [0.461,2.698] [0.602,2.757] [-1.274,1.365]	17.8 21.3 22.5 24.3 22.6
IX. Common A1	16	2891.2	560	1771.2	11.29	1	p <0.001	0.465 [0.193,0.737]	18.5 20.8 22.8 24.2 24.4
X. Common A2	16	2902.01	560	1782.01	0.47	1	0.49	0.123 [-0.231,0.477]	16.9 19.0 20.4 22.2 23.2

Table 19: Effect of GRSS on common genetic factors and mean BMI by sex in a subsample of genotyped ABD twin pairs (Method 2)

Table: FEMALES IP2f

Model	EP	-2LL	df	AIC	Diff LL	Difdf	p-value
I. Full	47	4780.75	1045	2690.75	-	-	-
I.a Drop Cfactor, Efactor	34	4791.71	1058	2675.71	10.96	13	0.61
I.b Reduced to best-fit from full sample	29	4794.87	1063	2668.87 II.	14.12 3.16	18 5	0.72 0.68
II. Add PCs	39	4780.91	1053	2674.91	-	-	-
III. GRSS Age 8 beta1 0.194 [-0.346,0.726]	40	4780.4	1052	2676.4	0.52	1	0.47
IV. GRSS Age 11 beta2 -0.018 [-0.545,0.516]	40	4780.91	1052	2676.91	<0.001	1	0.95
V. GRSS Age 13 beta3 0.486 [-0.061,1.039]	40	4777.85	1052	2673.85	3.06	1	0.08
VI. GRSS Age 15 beta4 -0.550 [-1.121,0.005]	40	4777.11	1052	2673.11	3.8	1	0.05
VII. GRSS Age 17 beta5 -0.390 [-0.975,0.175]	40	4777.72	1052	2673.72	3.19	1	0.07
VIII. GRSS Age 8-18 0.615 0.560 0.900 0.143 0.797 [-0.122,1.354] [-0.236,1.356] [0.023,1.781] [-0.76,1.048] [-0.398,1.983]	44	4772.51	1048	2676.51	8.4	5	0.14
IX. GRSS Af1 betaAf1 0.204 [-0.027,0.439]	40	4777.9	1052	2673.9	3.01	1	0.08
X. GRSS Af2 betaAf2 0.006 [-0.304,0.317]	40	4780.91	1052	2676.91	<0.001	1	0.97

Model	ac11 stdPathEst stdPath^2 [95%CI]	ac21 stdPathEst stdPath^2 [95%CI]	ac31 stdPathEst stdPath^2 [95%CI]	ac41 stdPathEst stdPath^2 [95%CI]	ac51 stdPathEst stdPath^2 [95%CI]	ac22 stdPathEst stdPath^2 [95%CI]	ac32 stdPathEst stdPath^2 [95%CI]	ac42 stdPathEst stdPath^2 [95%CI]	ac52 stdPathEst stdPath^2 [95%CI]
I. Full	2.469 0.835 0.696 [0.339,0.899]	3.226 0.909 0.827 [0.504,0.917]	3.576 0.851 0.724 [0.42,0.872]	3.243 0.769 0.591 [0.285,0.83]	2.92 0.609 0.371 [0.094,0.694]	-0.38 -0.107 0.011 [0,0.094]	1.062 0.253 0.064 [0,0.228]	1.973 0.468 0.219 [0.05,0.431]	2.584 0.539 0.29 [0.043,0.522]
I.a	2.582 0.873 0.761 [0.369,0.906]	3.34 0.941 0.885 [0.524,0.921]	3.482 0.829 0.688 [0.424,0.793]	3.296 0.782 0.611 [0.276,0.794]	3.067 0.641 0.411 [0.086,0.664]	-0.253 -0.071 0.005 [0,0.072]	1.144 0.272 0.074 [0.004,0.233]	2.216 0.526 0.277 [0.099,0.463]	2.853 0.596 0.356 [0.057,0.557]
I.b	2.745 0.932 0.868 [0.808,0.909]	3.135 0.883 0.78 [0.677,0.853]	3.326 0.788 0.62 [0.494,0.724]	3.056 0.721 0.52 [0.369,0.657]	3.201 0.664 0.441 [0.254,0.635]	-	-1.221 -0.289 0.084 [0.019,0.179]	-2.294 -0.541 0.293 [0.172,0.422]	-3.161 -0.656 0.43 [0.24,0.617]
II.	2.707 0.93 0.864 [0.802,0.906]	3.098 0.881 0.776 [0.671,0.85]	3.247 0.783 0.614 [0.486,0.719]	2.977 0.71 0.505 [0.351,0.644]	3.111 0.654 0.428 [0.237,0.631]	-	1.185 0.286 0.082 [0.017,0.179]	2.3 0.549 0.301 [0.179,0.433]	3.153 0.663 0.44 [0.24,0.631]
III.	2.699 0.929 0.864 [0.801,0.906]	3.101 0.883 0.779 [0.675,0.853]	3.24 0.781 0.611 [0.484,0.716]	2.975 0.711 0.505 [0.353,0.644]	3.079 0.647 0.419 [0.232,0.622]	-	1.201 0.29 0.084 [0.019,0.182]	2.304 0.55 0.303 [0.18,0.433]	3.186 0.67 0.449 [0.248,0.636]
IV.	2.707 0.93 0.864 [0.802,0.906]	3.099 0.881 0.777 [0.671,0.851]	3.248 0.784 0.614 [0.486,0.719]	2.978 0.711 0.505 [0.351,0.644]	3.113 0.655 0.428 [0.237,0.632]	-	1.184 0.286 0.082 [0.017,0.18]	2.299 0.549 0.301 [0.178,0.433]	3.151 0.663 0.439 [0.239,0.631]
V.	2.709 0.93 0.864 [0.803,0.906]	3.089 0.879 0.773 [0.666,0.849]	3.248 0.785 0.616 [0.488,0.721]	2.992 0.715 0.511 [0.359,0.648]	3.167 0.667 0.445 [0.247,0.645]	-	1.134 0.274 0.075 [0.013,0.172]	2.258 0.54 0.291 [0.171,0.423]	3.081 0.649 0.421 [0.226,0.62]
VI.	2.7 0.929 0.864 [0.802,0.906]	3.083 0.879 0.772 [0.665,0.847]	3.249 0.781 0.61 [0.481,0.716]	2.969 0.708 0.502 [0.347,0.641]	3.112 0.657 0.432 [0.234,0.641]	-	1.219 0.293 0.086 [0.019,0.186]	2.304 0.55 0.302 [0.178,0.435]	3.12 0.659 0.434 [0.229,0.633]
VII.	2.697 0.929 0.864 [0.802,0.906]	3.079 0.878 0.77 [0.663,0.847]	3.239 0.779 0.608 [0.477,0.715]	2.948 0.703 0.495 [0.339,0.637]	3.105 0.659 0.434 [0.23,0.65]	-	1.232 0.296 0.088 [0.02,0.189]	2.31 0.551 0.304 [0.176,0.438]	3.096 0.657 0.431 [0.218,0.636]
VIII.	2.676 0.928 0.862 [0.799,0.905]	3.058 0.878 0.771 [0.663,0.847]	3.214 0.777 0.604 [0.474,0.712]	2.925 0.702 0.493 [0.338,0.635]	3.072 0.653 0.427 [0.228,0.64]	-	1.218 0.295 0.087 [0.019,0.189]	2.299 0.552 0.305 [0.178,0.439]	3.108 0.661 0.437 [0.227,0.637]
IX.	2.68 0.928 0.862 [0.8,0.905]	3.062 0.878 0.771 [0.666,0.847]	3.217 0.779 0.607 [0.477,0.714]	2.907 0.698 0.487 [0.33,0.63]	3.086 0.653 0.426 [0.234,0.627]	-	1.209 0.293 0.086 [0.019,0.185]	2.326 0.558 0.312 [0.186,0.447]	3.135 0.663 0.44 [0.242,0.632]
X.	2.707 0.93 0.864 [0.802,0.906]	3.098 0.881 0.776 [0.671,0.851]	3.246 0.783 0.614 [0.485,0.719]	2.976 0.71 0.505 [0.351,0.644]	3.109 0.654 0.428 [0.237,0.632]	-	1.187 0.286 0.082 [0.017,0.18]	2.3 0.549 0.302 [0.178,0.433]	3.154 0.663 0.44 [0.239,0.631]

Model	cc11 stdPathEst stdPath^2 [95%CI]	cc21 stdPathEst stdPath^2 [95%CI]	cc31 stdPathEst stdPath^2 [95%CI]	cc41 stdPathEst stdPath^2 [95%CI]	cc51 stdPathEst stdPath^2 [95%CI]	cc22 stdPathEst stdPath^2 [95%CI]	cc32 stdPathEst stdPath^2 [95%CI]	cc42 stdPathEst stdPath^2 [95%CI]	cc52 stdPathEst stdPath^2 [95%CI]
I. Full	1.086 0.367 0.135 [0,0.496]	0.506 0.143 0.02 [0,0.322]	0.437 0.104 0.011 [0,0.267]	0.663 0.157 0.025 [0,0.311]	2.06 0.429 0.184 [0,0.551]	-0.61 -0.172 0.03 [0,0.18]	-0.111 -0.026 0.001 [0,0.165]	-0.977 -0.232 0.054 [0,0.223]	-0.664 -0.139 0.019 [0,0.287]
I.a	-0.961 -0.325 0.106 [0,0.488]	-0.086 -0.024 0.001 [0,0.334]	-0.251 -0.06 0.004 [0,0.257]	-0.128 -0.03 0.001 [0,0.319]	-1.541 -0.322 0.104 [0,0.575]	-	-	-	-
I.b	-	1.009 0.284 0.081 [0.005,0.189]	1.095 0.259 0.067 [0.007,0.162]	1.16 0.274 0.075 [0.001,0.138]	-	-	-	-	-
II.	-	0.962 0.274 0.075 [0.004,0.187]	1.103 0.266 0.071 [0.007,0.174]	1.175 0.28 0.079 [0.001,0.143]	-	-	-	-	-
III.	-	0.947 0.27 0.073 [0.003,0.184]	1.11 0.268 0.072 [0.007,0.177]	1.159 0.277 0.077 [0.001,0.142]	-	-	-	-	-
IV.	-	0.96 0.273 0.074 [0.003,0.188]	1.104 0.266 0.071 [0.007,0.174]	1.176 0.281 0.079 [0.001,0.143]	-	-	-	-	-
V.	-	0.891 0.254 0.064 [0.002,0.18]	1.174 0.284 0.08 [0.009,0.193]	1.194 0.285 0.081 [0.003,0.146]	-	-	-	-	-
VI.	-	0.953 0.272 0.074 [0.004,0.192]	1.168 0.281 0.079 [0.011,0.184]	1.202 0.287 0.082 [0.003,0.147]	-	-	-	-	-
VII.	-	-0.985 -0.281 0.079 [0.005,0.197]	-1.164 -0.28 0.078 [0.011,0.182]	-1.234 -0.294 0.087 [0.005,0.153]	-	-	-	-	-
VIII.	-	0.936 0.269 0.072 [0.003,0.191]	1.18 0.285 0.081 [0.012,0.191]	1.212 0.291 0.085 [0.004,0.151]	-	-	-	-	-
IX.	-	0.996 0.286 0.082 [0.006,0.194]	1.115 0.27 0.073 [0.009,0.174]	1.204 0.289 0.084 [0.004,0.148]	-	-	-	-	-
X.	-	0.963 0.274 0.075 [0.004,0.187]	1.102 0.266 0.071 [0.006,0.174]	1.175 0.281 0.079 [0.001,0.143]	-	-	-	-	-

Model	ec11 stdPathEst stdPath^2 [95%CI]	ec21 stdPathEst stdPath^2 [95%CI]	ec31 stdPathEst stdPath^2 [95%CI]	ec41 stdPathEst stdPath^2 [95%CI]	ec51 stdPathEst stdPath^2 [95%CI]	ec22 stdPathEst stdPath^2 [95%CI]	ec32 stdPathEst stdPath^2 [95%CI]	ec42 stdPathEst stdPath^2 [95%CI]	ec52 stdPathEst stdPath^2 [95%CI]
I. Full	-0.756 -0.255 0.065 [0.031,0.19]	-1.055 -0.297 0.088 [0.021,0.143]	-1.01 -0.24 0.058 [0.011,0.1]	-0.938 -0.223 0.05 [0.007,0.091]	-1.087 -0.227 0.051 [0,0.109]	-0.543 -0.153 0.023 [0,0.064]	0.516 0.123 0.015 [0,0.077]	0.632 0.15 0.022 [0.002,0.109]	1.022 0.213 0.045 [0.004,0.185]
I.a	-0.69 -0.233 0.054 [0.022,0.106]	-0.687 -0.194 0.037 [0.015,0.071]	-1.157 -0.276 0.076 [0.045,0.119]	-1.099 -0.261 0.068 [0.04,0.108]	-1.349 -0.282 0.08 [0.044,0.131]	-	-	-	-
I.b	-0.696 -0.236 0.056 [0.023,0.106]	-0.682 -0.192 0.037 [0.014,0.072]	-1.168 -0.277 0.076 [0.045,0.12]	-1.094 -0.258 0.067 [0.039,0.107]	-1.345 -0.279 0.078 [0.042,0.129]	-	-	-	-
II.	0.69 0.237 0.056 [0.023,0.107]	0.683 0.194 0.038 [0.015,0.073]	1.152 0.278 0.077 [0.046,0.121]	1.136 0.271 0.074 [0.043,0.116]	1.343 0.282 0.08 [0.043,0.132]	-	-	-	-
III.	0.689 0.237 0.056 [0.023,0.107]	0.683 0.195 0.038 [0.015,0.073]	1.153 0.278 0.077 [0.046,0.121]	1.135 0.271 0.074 [0.043,0.117]	1.343 0.282 0.08 [0.043,0.132]	-	-	-	-
IV.	0.69 0.237 0.056 [0.023,0.107]	0.683 0.194 0.038 [0.015,0.073]	1.152 0.278 0.077 [0.046,0.121]	1.136 0.271 0.074 [0.043,0.116]	1.343 0.282 0.08 [0.043,0.132]	-	-	-	-
V.	-0.689 -0.236 0.056 [0.023,0.107]	-0.682 -0.194 0.038 [0.015,0.073]	-1.158 -0.28 0.078 [0.047,0.122]	-1.143 -0.273 0.075 [0.044,0.118]	-1.339 -0.282 0.08 [0.043,0.132]	-	-	-	-
VI.	0.69 0.238 0.056 [0.023,0.107]	0.682 0.194 0.038 [0.015,0.073]	1.156 0.278 0.077 [0.046,0.12]	1.135 0.271 0.073 [0.043,0.116]	1.34 0.283 0.08 [0.043,0.132]	-	-	-	-
VII.	0.692 0.238 0.057 [0.024,0.108]	0.683 0.195 0.038 [0.015,0.073]	1.153 0.278 0.077 [0.046,0.12]	1.135 0.271 0.073 [0.043,0.116]	1.341 0.284 0.081 [0.044,0.134]	-	-	-	-
VIII.	0.689 0.239 0.057 [0.024,0.109]	0.683 0.196 0.038 [0.015,0.074]	1.162 0.281 0.079 [0.048,0.122]	1.143 0.274 0.075 [0.045,0.119]	1.343 0.286 0.082 [0.044,0.135]	-	-	-	-
IX.	0.689 0.239 0.057 [0.023,0.109]	0.682 0.196 0.038 [0.015,0.074]	1.151 0.279 0.078 [0.047,0.121]	1.15 0.276 0.076 [0.045,0.12]	1.344 0.284 0.081 [0.044,0.133]	-	-	-	-
X.	0.69 0.237 0.056 [0.023,0.107]	0.683 0.194 0.038 [0.015,0.073]	1.152 0.278 0.077 [0.046,0.121]	1.136 0.271 0.074 [0.043,0.117]	1.343 0.282 0.08 [0.043,0.132]	-	-	-	-

Model	as11 stdPathEst stdPath^2 [95%CI]	as22 stdPathEst stdPath^2 [95%CI]	as33 stdPathEst stdPath^2 [95%CI]	as44 stdPathEst stdPath^2 [95%CI]	as55 stdPathEst stdPath^2 [95%CI]	cs11 stdPathEst stdPath^2 [95%CI]	cs22 stdPathEst stdPath^2 [95%CI]	cs33 stdPathEst stdPath^2 [95%CI]	cs44 stdPathEst stdPath^2 [95%CI]	cs55 stdPathEst stdPath^2 [95%CI]
I. Full	0.529 0.179 0.032 [0,0.163]	0 0 0 [0,0.142]	1.305 0.31 0.096 [0,0.183]	0 0 0 [0,0.07]	0 0 0 [0,0.14]	0 0 0 [0,0.122]	0 0 0 [0,0.045]	0 0 0 [0,0.094]	0 0 0 [0,0.053]	0 0 0 [0,0.134]
I.a	0 0 0 [0,0.152]	0 0 0 [0,0.141]	-1.506 -0.359 0.129 [0,0.189]	0 0 0 [0,0.079]	0 0 0 [0,0.146]	-	-	-	-	-
I.b	-	-0.603 -0.17 0.029 [0,0.131]	1.485 0.352 0.124 [0.045,0.186]	-0.149 -0.035 0.001 [0,0.087]	-	-	-	-	-	-
II.	-	0.68 0.193 0.037 [0,0.141]	1.48 0.357 0.128 [0.042,0.192]	0 0 0 [0,0.09]	-	-	-	-	-	-
III.	-	0.671 0.191 0.037 [0,0.138]	1.483 0.358 0.128 [0.039,0.193]	0.056 0.013 0 [0,0.09]	-	-	-	-	-	-
IV.	-	-0.682 -0.194 0.038 [0,0.141]	1.48 0.357 0.128 [0.041,0.192]	0 0 0 [0,0.09]	-	-	-	-	-	-
V.	-	0.79 0.225 0.051 [0,0.152]	1.447 0.35 0.122 [0.031,0.188]	0 0 0 [0,0.088]	-	-	-	-	-	-
VI.	-	0.726 0.207 0.043 [0,0.147]	1.441 0.346 0.12 [0.034,0.183]	0 0 0 [0,0.089]	-	-	-	-	-	-
VII.	-	0.692 0.197 0.039 [0,0.145]	1.445 0.348 0.121 [0.038,0.184]	0 0 0 [0,0.093]	-	-	-	-	-	-
VIII.	-	0.725 0.208 0.043 [0,0.147]	1.437 0.348 0.121 [0.03,0.185]	0 0 0 [0,0.093]	-	-	-	-	-	-
IX.	-	0.637 0.183 0.033 [0,0.139]	1.478 0.358 0.128 [0.048,0.192]	0 0 0 [0,0.089]	-	-	-	-	-	-
X.	-	0.678 0.193 0.037 [0,0.141]	1.481 0.357 0.128 [0.042,0.192]	0 0 0 [0,0.091]	-	-	-	-	-	-

Model	es11 stdPathEst stdPath^2 [95%CI]	es22 stdPathEst stdPath^2 [95%CI]	es33 stdPathEst stdPath^2 [95%CI]	es44 stdPathEst stdPath^2 [95%CI]	e55 stdPathEst stdPath^2 [95%CI]
I. Full	0.792 0.268 0.072 [0,0.114]	0 0 0 [0,0.095]	0.75 0.178 0.032 [0.011,0.06]	0.839 0.199 0.04 [0,0.066]	-0.945 -0.197 0.039 [0,0.094]
I.a	0.831 0.281 0.079 [0.047,0.126]	-0.952 -0.268 0.072 [0.045,0.113]	-0.724 -0.172 0.03 [0.01,0.053]	0.875 0.208 0.043 [0.021,0.068]	1.066 0.223 0.05 [0.024,0.086]
I.b	-0.815 -0.277 0.077 [0.046,0.123]	-0.962 -0.271 0.074 [0.046,0.117]	0.715 0.169 0.029 [0.008,0.052]	-0.891 -0.21 0.044 [0.022,0.07]	1.082 0.224 0.05 [0.024,0.087]
II.	0.822 0.282 0.08 [0.048,0.127]	0.955 0.272 0.074 [0.046,0.118]	0.701 0.169 0.029 [0.009,0.051]	0.855 0.204 0.042 [0.019,0.068]	1.093 0.23 0.053 [0.027,0.09]
III.	0.823 0.283 0.08 [0.049,0.128]	0.955 0.272 0.074 [0.046,0.118]	0.7 0.169 0.028 [0.009,0.051]	0.856 0.204 0.042 [0.019,0.068]	1.092 0.23 0.053 [0.027,0.09]
IV.	0.822 0.282 0.08 [0.048,0.127]	0.955 0.272 0.074 [0.046,0.118]	0.701 0.169 0.029 [0.009,0.051]	0.855 0.204 0.042 [0.019,0.068]	1.093 0.23 0.053 [0.027,0.09]
V.	0.823 0.282 0.08 [0.049,0.127]	-0.954 -0.272 0.074 [0.046,0.117]	-0.692 -0.167 0.028 [0.009,0.05]	-0.848 -0.203 0.041 [0.019,0.067]	1.099 0.232 0.054 [0.027,0.091]
VI.	0.821 0.282 0.08 [0.049,0.126]	0.954 0.272 0.074 [0.046,0.117]	0.695 0.167 0.028 [0.009,0.05]	0.85 0.203 0.041 [0.019,0.067]	1.096 0.231 0.054 [0.027,0.092]
VII.	0.819 0.282 0.08 [0.048,0.126]	0.954 0.272 0.074 [0.046,0.118]	0.699 0.168 0.028 [0.009,0.05]	0.851 0.203 0.041 [0.019,0.067]	1.097 0.233 0.054 [0.027,0.093]
VIII.	0.821 0.285 0.081 [0.05,0.128]	0.954 0.274 0.075 [0.047,0.119]	0.689 0.167 0.028 [0.009,0.05]	0.85 0.204 0.042 [0.019,0.067]	1.097 0.233 0.054 [0.027,0.093]
IX.	0.822 0.285 0.081 [0.049,0.128]	0.957 0.274 0.075 [0.047,0.12]	0.7 0.169 0.029 [0.009,0.051]	-0.85 -0.204 0.042 [0.019,0.068]	1.096 0.232 0.054 [0.028,0.093]
X.	0.822 0.282 0.08 [0.048,0.127]	0.955 0.272 0.074 [0.046,0.118]	0.701 0.169 0.029 [0.009,0.051]	0.855 0.204 0.042 [0.019,0.068]	1.093 0.23 0.053 [0.027,0.09]

Table: MALES IP2f

Model	EP	-2LL	df	AIC	Diff LL	Difdf	p-value
I. Full	47	2796.55	645	1506.55	-	-	-
I.a Drop Cfactor, Efactor	34	2801.54	658	1485.54	4.99	13	0.98
I.b Reduced to best-fit from full sample	29	2815.13	663	1489.13 II.	18.58 13.59	18 5	0.42 0.02
II. Add PCs	39	2787.218	653	1481.22	-	-	-
III. GRSS Age 8 beta1 -0.083 [-0.846,0.689]	40	2787.17	652	1483.17	0.05	1	0.83
IV. GRSS Age 11 beta2 0.334 [-0.372,1.045]	40	2786.33	652	1482.33	0.88	1	0.35
V. GRSS Age 13 beta3 0.445 [-0.181,1.087]	40	2785.25	652	1481.25	1.97	1	0.16
VI. GRSS Age 15 beta4 0.384 [-0.211,0.982]	40	2785.58	652	1481.58	1.63	1	0.2
VII. GRSS Age 17 beta5 0.538 [-0.062,1.137]	40	2787.02	652	1483.02	0.2	1	0.65
VIII. GRSS Age 8-18 0.923 2.054 2.442 2.271 1.870 [0.115,1.736] [1.055,3.082] [1.389,3.498] [1.208,3.348] [0.498,3.252]	44	2764.2	648	1468.2	23.01	5	<0.001
IX. GRSS Af1 betaAf1 0.736 [0.399,1.094]	40	2768.56	652	1464.56	18.66	1	<0.001
X. GRSS Af2 betaAf2 -1.135 [-2.097,-0.602]	40	2771.87	652	1467.87	15.34	1	<0.001

Model	ac11 stdPathEst stdPath^2 [95%CI]	ac21 stdPathEst stdPath^2 [95%CI]	ac31 stdPathEst stdPath^2 [95%CI]	ac41 stdPathEst stdPath^2 [95%CI]	ac51 stdPathEst stdPath^2 [95%CI]	ac22 stdPathEst stdPath^2 [95%CI]	ac32 stdPathEst stdPath^2 [95%CI]	ac42 stdPathEst stdPath^2 [95%CI]	ac52 stdPathEst stdPath^2 [95%CI]
I. Full	1.771 0.831 0.691 [0.293,0.913]	2.339 0.708 0.501 [0.131,0.888]	3.149 0.844 0.713 [0.34,0.907]	3.1 0.826 0.682 [0.311,0.942]	2.771 0.675 0.455 [0.109,0.899]	1.495 0.452 0.204 [0,0.557]	1.448 0.388 0.151 [0,0.52]	1.569 0.418 0.175 [0,0.523]	2.46 0.599 0.359 [0,0.724]
I.a	1.859 0.874 0.764 [0.406,0.916]	2.568 0.782 0.611 [0.274,0.886]	2.949 0.792 0.628 [0.324,0.885]	2.85 0.761 0.579 [0.287,0.887]	2.34 0.571 0.326 [0.086,0.773]	-1.038 -0.316 0.1 [0,0.456]	-1.81 -0.486 0.236 [0,0.539]	-1.951 -0.521 0.271 [0,0.534]	-2.942 -0.718 0.516 [0.001,0.755]
I.b	-1.519 -0.735 0.541 [0.204,0.957]	-1.766 -0.533 0.284 [0.017,0.882]	-2.428 -0.642 0.412 [0.081,0.922]	-2.809 -0.754 0.569 [0.082,0.947]	-3.517 -0.864 0.746 [0.002,0.931]	1.868 0.564 0.318 [0,0.627]	2.299 0.608 0.369 [0,0.673]	2.171 0.583 0.34 [0,0.85]	1.552 0.381 0.145 [0,0.913]
II.	1.974 0.935 0.875 [0.704,0.93]	2.376 0.715 0.511 [0.243,0.786]	2.78 0.739 0.546 [0.311,0.882]	3.045 0.813 0.661 [0.399,0.857]	2.544 0.619 0.383 [0.146,0.764]	1.017 0.306 0.094 [0,0.32]	1.715 0.456 0.208 [0,0.437]	1.973 0.527 0.278 [0.039,0.538]	2.964 0.721 0.52 [0.137,0.758]
III.	-1.992 -0.936 0.876 [0.698,0.932]	-2.389 -0.718 0.516 [0.241,0.794]	-2.791 -0.742 0.55 [0.307,0.891]	-3.055 -0.815 0.665 [0.396,0.871]	-2.561 -0.623 0.388 [0.143,0.802]	0.996 0.299 0.09 [0,0.324]	1.698 0.451 0.204 [0,0.441]	1.96 0.523 0.274 [0.022,0.542]	2.951 0.718 0.515 [0.099,0.761]
IV.	-1.977 -0.935 0.875 [0.703,0.93]	-2.344 -0.716 0.513 [0.243,0.787]	-2.799 -0.744 0.553 [0.317,0.889]	-3.048 -0.815 0.664 [0.404,0.87]	-2.555 -0.622 0.387 [0.15,0.795]	0.969 0.296 0.088 [0,0.313]	1.69 0.449 0.202 [0,0.43]	1.962 0.524 0.275 [0.024,0.533]	2.953 0.719 0.516 [0.106,0.754]
V.	-2 0.937 0.877 [0.709,0.932]	-2.413 -0.724 0.524 [0.26,0.79]	-2.715 -0.736 0.542 [0.311,0.876]	-3.08 -0.825 0.68 [0.422,0.86]	-2.594 -0.635 0.403 [0.162,0.771]	0.952 0.286 0.082 [0,0.303]	1.652 0.448 0.201 [0,0.427]	1.898 0.508 0.258 [0.035,0.515]	2.889 0.707 0.499 [0.128,0.741]
VI.	-1.965 -0.934 0.872 [0.666,0.93]	-2.314 -0.699 0.488 [0.222,0.786]	-2.732 -0.726 0.526 [0.293,0.889]	-2.971 -0.803 0.645 [0.384,0.807]	-2.504 -0.608 0.369 [0.137,0.798]	1.068 0.322 0.104 [0,0.329]	1.79 0.475 0.226 [0,0.453]	2.001 0.541 0.292 [0.027,0.552]	3.013 0.731 0.534 [0.104,0.768]
VII.	1.956 0.933 0.871 [0.668,0.929]	2.306 0.697 0.486 [0.217,0.778]	2.743 0.728 0.53 [0.293,0.884]	2.935 0.796 0.634 [0.369,0.801]	2.424 0.594 0.353 [0.123,0.747]	1.065 0.322 0.104 [0,0.33]	1.766 0.469 0.22 [0,0.449]	2.029 0.55 0.303 [0.048,0.566]	3.026 0.741 0.549 [0.151,0.781]
VIII.	1.913 0.931 0.866 [0.671,0.926]	2.196 0.694 0.482 [0.215,0.777]	2.523 0.712 0.507 [0.271,0.874]	2.848 0.799 0.639 [0.376,0.801]	2.377 0.597 0.356 [0.124,0.756]	0.946 0.299 0.089 [0,0.312]	1.668 0.471 0.222 [0,0.452]	1.93 0.542 0.293 [0.045,0.554]	2.929 0.735 0.541 [0.14,0.774]
IX.	1.805 0.914 0.835 [0.352,0.915]	2.352 0.736 0.541 [0.278,0.796]	2.733 0.758 0.575 [0.337,0.893]	2.893 0.806 0.65 [0.402,0.869]	2.421 0.607 0.368 [0.152,0.791]	0.796 0.249 0.062 [0,0.276]	1.495 0.415 0.172 [0,0.405]	1.91 0.532 0.283 [0.019,0.522]	2.902 0.727 0.529 [0.102,0.749]
X.	1.444 0.685 0.469 [0.215,0.732]	1.889 0.598 0.358 [0.111,0.573]	2.544 0.714 0.51 [0.234,0.714]	2.88 0.812 0.659 [0.302,0.843]	3.626 0.905 0.82 [0.546,0.917]	-1.52 -0.482 0.232 [0.065,0.498]	-1.788 -0.502 0.252 [0.079,0.507]	-1.745 -0.492 0.242 [0.057,0.623]	-1.115 -0.278 0.078 [0.001,0.346]

Model	cc11 stdPathEst stdPath^2 [95%CI]	cc21 stdPathEst stdPath^2 [95%CI]	cc31 stdPathEst stdPath^2 [95%CI]	cc41 stdPathEst stdPath^2 [95%CI]	cc51 stdPathEst stdPath^2 [95%CI]	cc22 stdPathEst stdPath^2 [95%CI]	cc32 stdPathEst stdPath^2 [95%CI]	cc42 stdPathEst stdPath^2 [95%CI]	cc52 stdPathEst stdPath^2 [95%CI]
I. Full	0.925 0.434 0.189 [0,0.555]	0.514 0.156 0.024 [0,0.361]	-0.354 -0.095 0.009 [0,0.165]	0.731 0.195 0.038 [0,0.24]	0.231 0.056 0.003 [0,0.235]	-1.283 -0.388 0.151 [0,0.352]	-0.203 -0.054 0.003 [0,0.145]	0.777 0.207 0.043 [0,0.199]	1.155 0.281 0.079 [0,0.331]
I.a	-0.7 -0.329 0.108 [0,0.463]	0.394 0.12 0.014 [0,0.26]	0.378 0.101 0.01 [0,0.143]	-1.094 -0.292 0.085 [0,0.278]	-0.945 -0.231 0.053 [0,0.264]	-	-	-	-
I.b	1.154 0.559 0.312 [0,0.618]	1.161 0.35 0.123 [0,0.335]	1.165 0.308 0.095 [0,0.188]	-	-	-	-	-	-
II.	-0.011 -0.005 0 [0,0.161]	-1.252 -0.377 0.142 [0,0.291]	-1.358 -0.361 0.13 [0.003,0.207]	-	-	-	-	-	-
III.	0.025 0.012 0 [0,0.173]	1.25 0.376 0.141 [0,0.291]	1.358 0.361 0.13 [0,0.207]	-	-	-	-	-	-
IV.	-0.002 -0.001 0 [0,0.326]	-1.211 -0.37 0.137 [0,0.289]	-1.357 -0.36 0.13 [0,0.207]	-	-	-	-	-	-
V.	0.015 0.007 0 [0,0.158]	1.253 0.376 0.141 [0,0.288]	1.365 0.37 0.137 [0.005,0.215]	-	-	-	-	-	-
VI.	-0.086 -0.041 0.002 [0,0.2]	-1.299 -0.392 0.154 [0,0.309]	-1.368 -0.363 0.132 [0,0.208]	-	-	-	-	-	-
VII.	-0.075 -0.036 0.001 [0,0.198]	-1.297 -0.392 0.154 [0,0.312]	-1.379 -0.366 0.134 [0.001,0.213]	-	-	-	-	-	-
VIII.	0.047 0.023 0.001 [0,0.192]	1.246 0.394 0.155 [0,0.316]	1.339 0.378 0.143 [0,0.224]	-	-	-	-	-	-
IX.	0.258 0.13 0.017 [0,0.493]	-1.136 -0.355 0.126 [0,0.283]	-1.294 -0.359 0.129 [0,0.208]	-	-	-	-	-	-
X.	-1.338 -0.635 0.403 [0.153,0.644]	-1.144 -0.362 0.131 [0.037,0.275]	-1.181 -0.332 0.11 [0.042,0.193]	-	-	-	-	-	-

Model	ec11 stdPathEst stdPath^2 [95%CI]	ec21 stdPathEst stdPath^2 [95%CI]	ec31 stdPathEst stdPath^2 [95%CI]	ec41 stdPathEst stdPath^2 [95%CI]	ec51 stdPathEst stdPath^2 [95%CI]	ec22 stdPathEst stdPath^2 [95%CI]	ec32 stdPathEst stdPath^2 [95%CI]	ec42 stdPathEst stdPath^2 [95%CI]	ec52 stdPathEst stdPath^2 [95%CI]
I. Full	0.738 0.347 0.12 [0.006,0.219]	0.405 0.123 0.015 [0,0.153]	0.644 0.173 0.03 [0.003,0.173]	0.128 0.034 0.001 [0,0.062]	0.34 0.083 0.007 [0,0.128]	0.876 0.265 0.07 [0,0.14]	0.831 0.223 0.05 [0,0.101]	0.757 0.202 0.041 [0,0.089]	0.823 0.2 0.04 [0,0.147]
I.a	0.4 0.188 0.035 [0.004,0.105]	0.956 0.291 0.085 [0.042,0.152]	1.004 0.27 0.073 [0.038,0.125]	0.743 0.198 0.039 [0.02,0.069]	0.916 0.224 0.05 [0.018,0.105]	-	-	-	-
I.b	0.412 0.199 0.04 [0,0.128]	0.962 0.29 0.084 [0.041,0.24]	1.009 0.267 0.071 [0.036,0.193]	0.733 0.197 0.039 [0.007,0.069]	0.899 0.221 0.049 [0.015,0.182]	-	-	-	-
II.	-0.364 -0.172 0.03 [0.002,0.096]	-0.868 -0.261 0.068 [0.031,0.129]	-0.961 -0.255 0.065 [0.033,0.115]	-0.712 -0.19 0.036 [0.017,0.065]	-0.822 -0.2 0.04 [0.012,0.089]	-	-	-	-
III.	0.365 0.171 0.029 [0.002,0.096]	0.868 0.261 0.068 [0.031,0.129]	0.961 0.255 0.065 [0.033,0.115]	0.712 0.19 0.036 [0.017,0.065]	0.822 0.2 0.04 [0.012,0.089]	-	-	-	-
IV.	0.367 0.174 0.03 [0.002,0.096]	0.866 0.265 0.07 [0.032,0.133]	0.962 0.256 0.065 [0.033,0.115]	0.711 0.19 0.036 [0.017,0.065]	0.823 0.2 0.04 [0.012,0.09]	-	-	-	-
V.	0.364 0.17 0.029 [0.001,0.094]	0.869 0.261 0.068 [0.031,0.129]	0.958 0.26 0.068 [0.034,0.119]	0.711 0.19 0.036 [0.017,0.065]	0.825 0.202 0.041 [0.012,0.091]	-	-	-	-
VI.	-0.364 -0.173 0.03 [0.002,0.097]	-0.868 -0.262 0.069 [0.031,0.13]	-0.961 -0.255 0.065 [0.033,0.114]	-0.713 -0.193 0.037 [0.017,0.067]	-0.821 -0.199 0.04 [0.012,0.089]	-	-	-	-
VII.	0.365 0.174 0.03 [0.002,0.098]	0.868 0.263 0.069 [0.031,0.13]	0.962 0.255 0.065 [0.033,0.115]	0.715 0.194 0.038 [0.017,0.067]	0.818 0.2 0.04 [0.012,0.09]	-	-	-	-
VIII.	0.373 0.181 0.033 [0.002,0.104]	0.864 0.273 0.075 [0.034,0.14]	0.952 0.269 0.072 [0.036,0.126]	0.709 0.199 0.04 [0.018,0.071]	0.825 0.207 0.043 [0.013,0.095]	-	-	-	-
IX.	0.368 0.186 0.035 [0.002,0.115]	0.861 0.269 0.073 [0.033,0.137]	0.951 0.264 0.07 [0.035,0.123]	0.703 0.196 0.038 [0.017,0.07]	0.831 0.208 0.043 [0.013,0.097]	-	-	-	-
X.	0.385 0.183 0.033 [0.003,0.103]	0.866 0.274 0.075 [0.034,0.14]	0.946 0.266 0.071 [0.035,0.124]	0.713 0.201 0.04 [0.018,0.073]	0.805 0.201 0.04 [0.011,0.093]	-	-	-	-

Model	as11 stdPathEst stdPath^2 [95%CI]	as22 stdPathEst stdPath^2 [95%CI]	as33 stdPathEst stdPath^2 [95%CI]	as44 stdPathEst stdPath^2 [95%CI]	as55 stdPathEst stdPath^2 [95%CI]	cs11 stdPathEst stdPath^2 [95%CI]	cs22 stdPathEst stdPath^2 [95%CI]	cs33 stdPathEst stdPath^2 [95%CI]	cs44 stdPathEst stdPath^2 [95%CI]	cs55 stdPathEst stdPath^2 [95%CI]
I. Full	0 0 0 [0,0.402]	0 0 0 [0,0.212]	0 0 0 [0,0.069]	0 0 0 [0,0.05]	0 0 0 [0,0.111]	0 0 0 [0,0.356]	0 0 0 [0,0.257]	0 0 0 [0,0.12]	0 0 0 [0,0.074]	0 0 0 [0,0.103]
I.a	0 0 0 [0,0.413]	-1.302 -0.396 0.157 [0,0.273]	0 0 0 [0,0.102]	0 0 0 [0,0.043]	0 0 0 [0,0.107]	-	-	-	-	-
I.b	-	-1.329 -0.401 0.161 [0,0.273]	-	0.601 0.161 0.026 [0,0.139]	-	-	-	-	-	-
II.	-	1.304 0.392 0.154 [0.076,0.267]	-	0 0 0 [0,0.067]	-	-	-	-	-	-
III.	-	-1.301 -0.391 0.153 [0.074,0.266]	-	0 0 0 [0,0.069]	-	-	-	-	-	-
IV.	-	1.309 0.4 0.16 [0.079,0.277]	-	0 0 0 [0,0.07]	-	-	-	-	-	-
V.	-	1.31 0.393 0.154 [0.069,0.27]	-	0 0 0 [0,0.072]	-	-	-	-	-	-
VI.	-	1.299 0.392 0.154 [0.076,0.268]	-	0 0 0 [0,0.072]	-	-	-	-	-	-
VII.	-	1.305 0.394 0.156 [0.08,0.269]	-	0 0 0 [0,0.067]	-	-	-	-	-	-
VIII.	-	1.282 0.405 0.164 [0.076,0.284]	-	0 0 0 [0,0.082]	-	-	-	-	-	-
IX.	-	1.294 0.405 0.164 [0.075,0.283]	-	0 0 0 [0,0.092]	-	-	-	-	-	-
X.	-	1.299 0.412 0.169 [0.093,0.28]	-	0.619 0.174 0.03 [0,0.099]	-	-	-	-	-	-

Model	es11 stdPathEst stdPath^2 [95%CI]	es22 stdPathEst stdPath^2 [95%CI]	es33 stdPathEst stdPath^2 [95%CI]	es44 stdPathEst stdPath^2 [95%CI]	e55 stdPathEst stdPath^2 [95%CI]
I. Full	0 0 0 [0,0.164]	-0.617 -0.187 0.035 [0,0.077]	0.788 0.211 0.045 [0,0.08]	-0.539 -0.144 0.021 [0,0.045]	0.983 0.239 0.057 [0,0.1]
I.a	-0.647 -0.304 0.093 [0.047,0.181]	0.597 0.182 0.033 [0.006,0.07]	-0.853 -0.229 0.053 [0.03,0.086]	0.594 0.159 0.025 [0.012,0.045]	0.96 0.234 0.055 [0.027,0.095]
I.b	0.678 0.328 0.108 [0.025,0.222]	0.583 0.176 0.031 [0,0.068]	0.87 0.23 0.053 [0.029,0.13]	-0.604 -0.162 0.026 [0.012,0.072]	0.992 0.244 0.059 [0.029,0.105]
II.	0.653 0.309 0.096 [0.049,0.183]	0.591 0.178 0.032 [0.006,0.066]	0.845 0.225 0.05 [0.027,0.084]	0.595 0.159 0.025 [0.011,0.045]	0.981 0.239 0.057 [0.029,0.098]
III.	-0.653 -0.307 0.094 [0.048,0.184]	0.591 0.178 0.032 [0.006,0.066]	0.845 0.225 0.05 [0.027,0.084]	-0.595 -0.159 0.025 [0.011,0.045]	0.981 0.239 0.057 [0.029,0.098]
IV.	0.651 0.308 0.095 [0.049,0.182]	0.591 0.181 0.033 [0.006,0.069]	0.845 0.224 0.05 [0.027,0.084]	0.597 0.159 0.025 [0.011,0.046]	-0.98 -0.239 0.057 [0.029,0.098]
V.	0.654 0.306 0.094 [0.048,0.18]	-0.592 -0.177 0.031 [0.006,0.067]	-0.846 -0.229 0.053 [0.028,0.088]	0.596 0.16 0.025 [0.011,0.046]	0.979 0.24 0.057 [0.029,0.098]
VI.	-0.654 -0.311 0.097 [0.05,0.185]	0.589 0.178 0.032 [0.006,0.067]	0.846 0.225 0.05 [0.027,0.084]	-0.594 -0.161 0.026 [0.011,0.047]	-0.981 -0.238 0.057 [0.029,0.097]
VII.	0.653 0.312 0.097 [0.05,0.186]	0.589 0.178 0.032 [0.006,0.067]	0.847 0.225 0.051 [0.027,0.084]	0.592 0.161 0.026 [0.011,0.047]	0.984 0.241 0.058 [0.03,0.099]
VIII.	0.65 0.316 0.1 [0.051,0.193]	0.589 0.186 0.035 [0.006,0.073]	0.845 0.238 0.057 [0.031,0.094]	0.597 0.168 0.028 [0.012,0.051]	0.979 0.246 0.06 [0.031,0.104]
IX.	0.664 0.336 0.113 [0.058,0.219]	0.589 0.184 0.034 [0.006,0.072]	0.845 0.234 0.055 [0.029,0.093]	0.604 0.168 0.028 [0.012,0.051]	0.977 0.245 0.06 [0.03,0.105]
X.	0.648 0.307 0.095 [0.048,0.185]	0.581 0.184 0.034 [0.005,0.071]	0.854 0.24 0.058 [0.031,0.095]	0.596 0.168 0.028 [0.012,0.052]	0.999 0.249 0.062 [0.032,0.107]

SUPPLEMENTAL MATERIAL

Table 20: ABD sample sizes by age and zygosity

ABD Sample	8-10	11-12	13-14	15-16	17-18	Total Families
ABD BMI Female (pairs/singles)	199/4	241/1	300/3	303/3	173/7	607
MZ	141/3	159/0	199/1	223/3	124/4	
DZ	58/1	82/1	101/2	80/0	49/3	
ABD BMI Male (pairs/singles)	151/3	213/1	255/2	268/4	170/5	495
MZ	87/1	132/1	165/2	170/2	120/1	
DZ	64/2	81/0	92/0	98/2	50/4	
ABD BMI Female GRSS T1 & T2	77/3	101/0	135/0	138/0	91/0	242/45
MZ (pairs/singles)	64/3	80/0	108/0	118/0	74/0	
DZ (pairs/singles)	13/0	21/0	27/0	20/0	17/0	
ABD BMI Male GRSS T1 & T2	41/2	59/0	87/0	94/0	63/0	152/46
MZ (pairs/singles)	35/0	48/0	68/0	72/0	54/0	
DZ (pairs/singles)	6/2	11/0	19/0	22/0	9/0	
ABD BMI Female independent	119	159	194	195	132	359
MZ	64	80	109	119	74	
DZ	30	44	49	41	27	
DZO	25	35	36	35	31	
ABD BMI Male independent	74	105	135	160	102	258
MZ	35	48	69	72	54	
DZ	24	29	41	47	23	
DZO	15	28	25	41	25	

Chapter 5: Comparisons of energy intake and energy expenditure in overweight and obese women with and without binge eating disorder

Adapted from:

1) Bartholome LT*, Peterson RE*, Raatz SK, Raymond NC. *Authors contributed equally to this work. A comparison of the accuracy of self-reported intake with measured intake of a laboratory overeating episode in overweight and obese women with and without binge eating disorder. *Eur J Nutr.* 2012 Feb 3.
2) Raymond NC, Peterson RE, Bartholome LT, Raatz SK, Jensen MD, Levine JA. Comparisons of Energy Intake and Energy Expenditure in Overweight and Obese Women with and Without Binge Eating Disorder. *Obesity*. 2012 Apr;20(4):765-72. Epub 2011 Oct 20.

ABSTRACT

The purpose of this study was to determine whether there are differences in energy intake or energy expenditure that distinguish overweight/obese women with and without binge eating disorder (BED). Furthermore, research has demonstrated significant underreporting of food intake in obese individuals with and without BED. An improved understanding of the accuracy of self-reported food intake is central to diagnosis of eating disorders and monitoring response to treatment. Seventeen overweight/obese women with BED and 17 overweight/obese controls completed random 24-hour dietary recall interviews, participated in a laboratory eating episode and had total daily energy expenditure (TDEE) assessed by the doubly labeled water technique with concurrent food log data collection. Results indicated no between group differences in TDEE, basal metabolic rate (BMR) or thermal effect of food (TEF). According to dietary recall data, the BED group had significantly higher caloric intake on days when they had binge eating episodes than on days when they did not (3255 vs. 2343 kilocalories (kcal)). There was no difference between BED non-binge day intake and control group intake (2233 vs. 2140 kcal). Similar results were found for food log data and laboratory measured intake. Furthermore, when comparing TDEE to dietary recall and food log data, both groups displayed significant underreporting of caloric intake of similar magnitudes ranging 20-33%. Predicted energy requirements estimated via the Harris-Benedict equation underestimated measured TDEE by 23-24%. The BED group self-reported 90% of the laboratory measured intake compared to 98% for the control group. Mean differences between the methods indicated that on average both groups under-reported intake, however the mean difference between methods was significantly greater in the BED group. Findings confirm that those with BED consume significantly more than controls during a laboratory binge and controls tended to be more accurate in recalling their intake 24 hours later. Our data suggest that increased energy intake reported by BED individuals is due to increased food consumption and not metabolic or reporting differences.

INTRODUCTION

Binge eating disorder (BED) is currently classified in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) (190) as a provisional diagnosis requiring further study to support its utility as an eating disorders diagnosis. Two central criteria describe binge eating in the DSM-IV: 1) "Eating in a discrete period of time an amount of food that is definitely larger than most individuals would eat during a similar period of time and under similar circumstances" and 2) "a sense of lack of control" (190). This study was designed to examine whether there are differences in energy intake or energy expenditure patterns, which distinguish those with binge eating disorder from typical overweight/obese controls. Differences in these biological and behavioral factors between groups may help to clarify whether BED is a distinct eating disorder from obesity by identifying metabolic and food intake differences between groups.

In both clinical and research settings, the food intake data necessary to determine if an individual fulfills the first criteria above are collected utilizing self-report techniques. Throughout medicine, there are concerns regarding the accuracy of self-reported data. Research has demonstrated significant underreporting of food intake in obese individuals with (276, 277) and without BED (278-280). For those with BED, precise measurements of energy intake are associated with additional challenges since eating episodes are often secretive and associated with feelings of embarrassment and guilt over how much one is eating (281-283). These characteristics of binge eating may influence accuracy of reporting.

Despite the challenges associated with precise, objective measurement of eating behaviors, laboratory studies have been utilized to study food intake in obese women with and without BED. Our group and others have measured food intake in the laboratory through the administration of a test meal to simulate a binge eating episode. Test meal composition has varied by laboratory and included liquid meals (284, 285), single item meals (286-289) and multiple item arrays of food (287, 290-293). Despite differing laboratory methodologies, results have consistently demonstrated that individuals with BED have greater total energy intake than non-BED weight matched controls when instructed to overeat. Furthermore, there is additional indirect evidence that women with BED are eating more than they expend as research has shown the proportion of women with BED in obese samples increases as BMI increases (294-296).

According to previous reports in the literature, having a positive energy balance (i.e., chronic overfeeding) leads to increased TDEE (297-299). TDEE is most accurately measured using doubly labeled water (DLW) method, which estimates TDEE within 4-5% in free living individuals (300). Reports utilizing the DLW method to measure TDEE suggest that obese individuals report approximately 60% of their actual energy intake (278, 279). However, few studies have specifically examined the accuracy of self-reported food intake data in BED. Yanovski's group examined the accuracy of self-reported data by comparing average daily food intake assessed by food records to estimated daily energy expenditure calculated by the Harris-Benedict equation (HBE) (276). The BED group reported energy intake equivalent to 94% of their predicted energy requirements compared to 60% in the non-BED obese group. The authors suggested that the BED group may be more accurate in reporting food intake than

controls because of the psychological distress associated with binge eating may make the experience more memorable and thereby more accurate. However, given the limited number of studies, it remains unclear whether those with BED and non-BED individuals report food intake with comparable levels of accuracy. If these groups do not have consistent reporting patterns, food intake data must be interpreted with caution when trying to determine if they manifest distinctive eating patterns. An improved understanding of the accuracy of self-reported food intake data is central to distinguishing BED from typical obesity, making sound diagnosis, and monitoring response to treatment.

In the current study, we sought to replicate findings by our group and others that participants with BED will consume more kilocalories than their non-BED counterparts when instructed to overeat in the laboratory. Secondly, we hypothesize that those with BED are in a constant state of positive energy balance and therefore will have an elevated TDEE compared to the non-binge eating women. This is the first study to utilize the DLW method in the assessment of TDEE in BED. Additionally, Measured food intake in the laboratory was compared to dietary recall data to ascertain the accuracy of participants' recall of the overeating episode. We hypothesized that the overweight/obese control group would report approximately 60% of measured test meal intake, consistent with previous reports (278, 279), while the BED group would be more accurate as observed by Yanovksi and colleagues (276). Further, we sought to confirm the positive correlation between total food intake and BMI in those with BED during an overeating episode. A final aim of this paper was to explore the possibility of reduced dietary intake as a potential precursor to binge eating in BED. We compared food intake preceding the laboratory overeating episode to test meal intake to ascertain whether caloric intake before the test meal influences eating during the test meal.

METHODS AND PROCEDURES

Participants

Participants were 17 women who met DSM IV criteria for BED as defined in the appendix titled *Criteria Sets and Axes Provided for Further Study* and 17 women with no history of eating disorder symptoms including binge eating or purging behaviors. In order to participate in the study, women were required to be between the ages of 18 and 55, have no history of substance abuse or dependence within the six months prior to the study, and have no unstable comorbid medical or psychiatric conditions. Participants could not be smokers, pregnant, nursing or on a weight reduction diet as all of these conditions affect energy metabolism. Because of the difficulty recruiting participants free of psychiatric medications, participants were not excluded if they were on a stable dose (for at least 6 months) of antidepressant medication, were psychiatrically stable and had no plans to modify their medication during the duration of the study. Six participants with BED and 2 controls were on antidepressants during the time they were participating in the study. The study was conducted at the General Clinical Research Center (GCRC) of the University of Minnesota. This protocol was reviewed and approved by the

Institutional Review Board at the University of Minnesota and all participants partook in the informed consent process and signed a consent form.

Recruitment was performed by newspaper advertisements inviting overweight women aged 18 to 55 years old to participate in a paid research study at the University of Minnesota. A telephone screen was used to assess preliminary eligibility for the BED and control groups. Participants meeting initial criteria were scheduled for a complete evaluation at the Ambulatory Research Center (ARC) to determine eligibility. During this evaluation participants were interviewed using the Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-I/P) (301); the Structured Clinical Interview for Axis II Personality Disorders (SCID-II) (302); and the Eating Disorder Examination, version 12.0D (EDE) (303). These assessments were used to confirm that BED participants fulfilled diagnostic criteria and to rule out any history of eating disorder symptoms in the control group. A physical exam, complete blood count, basic metabolic panel, and thyroid and liver function tests were performed to detect unknown medical conditions that could influence eligibility.

As part of the initial evaluation, participants were interviewed by a registered dietitian who was blind to their diagnostic status to assess typical food intake patterns, food selection, and preferred snack foods. Participants were presented with a standardized list of food items and asked to indicate which appealed to them. In addition, participants were asked if they had other favorite foods or recipes that they consumed when overeating. Based on this information, the dietitian created a snack tray personalized to each participant's eating preferences for the laboratory overeating episode. Snack trays included 6-10 food items, consisting of both savory and sweet, in quantities 2-3 times what participants reportedly consumed during an overeating episode.

Eligible participants were then scheduled for a 24-hour inpatient stay at the GCRC during which they would engage in a laboratory overeating episode and subsequently complete a telephone dietary recall of 24-hour period including the test meal. Patients were not informed that they were scheduled for a dietary recall interview until after completion of the overeating episode. This was done to ensure that knowledge of the recall would not influence eating behaviors in the laboratory. In addition to collecting food intake data for the test meal, the dietary recall protocol gathered self-reported food intake for the periods preceding and following the overeating episode. This enabled a comparison of pre-binge and post-binge food intake to intake during the test meal.

Eligible participants were scheduled for two procedures, six random 24-hour dietary recall interviews and a 24-hour inpatient stay on the General Clinical Research Center (GCRC). On the day of admission, participants were instructed not to eat after 12:00 noon. While on the inpatient unit, they consumed doubly labeled water for the TDEE measurement, received two baseline (DXA) scans and had BMR and TEF measured using indirect calorimetry. Details of each of these methods are provided below. Participants' height and weight were measured on admission. Weight was repeated two weeks later. The inpatient stay was scheduled to coincide with the luteal phase of the menstrual cycle, confirmed by estradiol and progesterone levels, to control for hormonal influences on food intake.

Laboratory binge eating episode

Participants were instructed not to consume any food or caloric beverages after 12 pm and to arrive at the GCRC no later than 5:30 pm. After admission procedures, participants were presented with a standard hospital dinner plus an excess of their preferential binge foods as ascertained by the dietary assessment. They were instructed to "Let yourself go and eat as much as you like". Participants were left alone to eat and told to notify the research team when they were finished with the meal. This same laboratory test meal protocol has been utilized in previous work by our group (293).

Upon completion of the meal, food trays were removed from the room. All food items presented to participants were weighed in the GCRC metabolic kitchen prior to service, and remaining portions were weighed after completion of the overeating episode. The exact quantity of each item consumed was calculated by difference in mass. The computer program Nutritionist IV (304) was used to calculate total food intake in kcal and grams (gm) and macronutrient intake in gm. To compare our results with others, macronutrient values in kcal were estimated from measured values in grams by the following standard conversion: 4.0 kcal/gm carbohydrate, 4.0 kcal/gm protein, and 9.0 kcal/gm fat.

Twenty-four hour dietary recall

Over a six to eight week period of time (that excluded the DLW data collection period) each participant received six random 24-hour dietary recall interviews that were conducted by the staff of the Nutrition Coordinating Center (NCC), Department of Epidemiology, School of Public Health, University of Minnesota. Four of the six were conducted during weekdays and two on weekends as this best approximates normal intake. The dietary recall interviews involved a detailed discussion of food intake and portion sizes with expert interviewers. The 24-hour dietary recall interview protocol has been described in previous studies by our group (305, 306). Dietary interviewers collected the 24-hour dietary recalls using a current version of the database each year. At the end of data collection, nutrients were recalculated for all dietary intake records on the most current version of the Nutrition Data System for Research (NDS-R) software version 4.01, Food and Nutrient Database 30, released November 1999. NDS-R is developed and maintained by the NCC, University of Minnesota, Minneapolis, MN. The NDS-R system prompts the interviewer to ask detailed questions about food intake over a 24-hour period. The interviewer asks the participant to recall the first eating episode during the 24-hour period. As the interviewer records food items during that eating episode the program prompts the interviewer to ask about additional foods that may be typically eaten with the specific item (e.g. condiments with hot dogs or the type of milk or sugar added to cereal). When the first eating episode is fully explored, the interviewer asks about the next eating episode and proceeds in this fashion through the entire 24 hour period. Prior to the data collection participants were trained in the use of food-portion visuals (picture of containers and shapes of specific quantities that are drawn to scale) to estimate dietary intake as described by Posner (307).

Additionally, on the afternoon following the laboratory test meal, participants completed a dietary recall interview for the 24-hour time period from midnight to midnight during which they engaged in the overeating episode. At the time the dietary

recall of the inpatient binge eating episode was collected, all participants had already competed at least one random recall with the NCC interviewers as part of the larger research protocol in which they were participating. Following collection of dietary recall data, eating episodes that occurred during the 24-hour period were defined as pre-binge, binge or post-binge. Pre-binge intake was defined as food consumption beginning at 12 am up to delivery of the test meal. Binge intake included only the test meal administered at the GCRC. Post-binge intake was defined as food consumption after the test meal until 11:59 pm. This breakdown enabled a comparison of pre-binge and binge intake to examine the role of reduced caloric intake as a precursor to binge eating episodes.

Food log

During the two weeks of urine collection participants also kept a food diary so that recorded intake could be compared to measured TDEE during the two week period of time. Participants were trained in the use of food logs by a training tape provided by the GCRC dietician. Food logs were routinely reviewed by the research team and further questions regarding intake were asked if recorded data lacked sufficient detail for calculation of energy intake.

Basal metabolic rate (BMR) and thermic effect of food (TEF)

BMR and TEF were measured using the Delta Track Metabolic Cart (SensorMedics, Yorba Linda, CA). BMR and TEF were collected for two participants (one BED and one control) on a SensorMedics Vmax 29 Metabolic cart (SensorMedics, Yorba Linda, CA) because of technical issues with equipment. Participants were awakened at a standardized morning hour, allowed to void, and then rested for one-half hour before BMR was measured. BMR was assessed using a thirty minute recording under the plastic hood while awake, in a semi-recumbent position in bed. The first 10 minutes were used to obtain a stable baseline. BMR was then calculated from the average of the next 20 minutes of data collection. Participants then drank a standardized oral meal replacement formula (Ensure High Protein, Abbot Laboratories) which contained 250 kilocalories (protein 14.4%, carbohydrate 64.0%, Fat 21.6%). TEF or postprandial thermogenesis was measured based on data collected over the next 5 hours by placing the participant under the hood to collect data for 15 minutes of every 30 minutes to prevent participant fatigue or agitation. The first 5 minutes of every 15 minute period was used to establish a stable baseline. Conventional methods were used to calculate daily TEF.

Total daily energy expenditure

TDEE was measured over 14 days using the doubly labeled water protocol (25, 30, 31). Baseline urine specimens were collected immediately prior to the timed ingestion of the isotopes (^{2}H and O^{18}). The amount administered was calculated according to a standardized procedure (25, 26). Following timed administration of the isotopes, urine samples were collected at 12 hour intervals each day for 14 days. Date and exact collection times were recorded on each bottle and specimens were dropped off to the clinic every three to four days during the two weeks of data collection. TDEE was

derived using the slope-intercept equations described by Coward, et al. (32). Validation studies have determined the precision of the method to be within 4-5% (33).

Assessment of change in body composition through repeated DXA scans

Two baseline DXA scans (Lunar Prodigy, General Electric Medical, Madison, WI) were collected on the day the DLW was administered. They were repeated two weeks later at the completion of DLW protocol. Assessment of body composition is essential because if body weight and composition are stable, energy intake must be equal to energy expenditure. Collection of body composition data allows for an accurate comparison of food intake data (collected via dietary recalls and food logs) to the TDEE measured by DLW. If there was no change in body weight or composition, the measured TDEE should be equal to energy intake. Therefore, by comparing reported food intake to measured energy expenditure, we examined the accuracy of food log data kept over the two-week period when TDEE was assessed.

Predicted energy requirements

The Harris-Benedict equation (HBE), commonly used in clinical settings, calculates resting metabolic rate based on gender, weight, height and age (34). Predicted energy requirements can be made by adjusting for activity level. To attain predicted energy needs, participants' HBE estimates were multiplied by 1.35 to account for light activity.

Analysis of DLW by isotope ratio/mass spectrometry

Deuterium and ^{18}O in urine were measured using a dual inlet ThermoFinnigan DeltaS Isotope Ratio Mass Spectrometer (ThermoFisher Scientific, Bremen, Germany). Deuterium was analyzed using an H-Device by reducing 1μL water via a chromium furnace held at 825°C. The deuterium produced was measured against a calibrated hydrogen reference gas. ^{18}O was measured in a separate assay by equilibration of urine with CO_2. 1ml urine was introduced into a 12ml exetainer and 5% CO_2 in Helium added to the tube. The sample was then allowed to equilibrate overnight at room temperature. Analysis of the $C^{18}O^{16}O$ produced was performed by measurement against a CO_2 reference gas using a breath bench carousel inlet. In both assays, calibration curves were prepared to which the samples were compared.

Statistical analysis

Descriptive statistics, Pearson correlation coefficients and analysis of variance were calculated using SPSS version 17.0. The reporting accuracy was defined by two methods. The first method was the *directional difference,* which was defined as the mean difference of measured intake minus reported intake. Negative values reflect over-reporting while positive values signify under-reporting. The second method to determine reporting accuracy was the *absolute difference*, which was defined as the absolute value of the measured intake less the reported intake. Greater absolute difference values indicated greater inaccuracy overall despite whether the difference arose from under or

over-reporting. Analysis of variance was used to determine between group differences on total and macronutrient intake, directional and absolute difference between laboratory and dietary recall, and energy consumption throughout the day. The proportion of energy intake from carbohydrate, fat and protein was examined by dividing the macronutrient intake by total intake. Pearson's correlation coefficient was used to determine the relationship between BMI and total food intake. Student's *t*-tests and paired samples correlation coefficients were used to compare within group differences on laboratory, dietary recall data and energy consumption throughout the day. To assess the difference between correlation coefficients between groups, Fischer's r-to-z transformations were used.

RESULTS

Demographic data

There were no statistically significant differences between groups with regard to age and BMI (Table 21). The BMI range for the participants was 25.6 to 51.9 with 20.7% of the sample overweight (4 BEDs and 2 controls). Baseline binge frequency according to EDE assessments in the BED group ranged from twice per week to daily with a group mean of 17 episodes per month (median = 12).

Metabolic measurements

There were no between group differences in TDEE, BMR, or TEF (Table 21). TDEE was significantly correlated with total food intake in kcal as assessed by 24-hour recall in the whole sample (n = 29, $r^2 = 0.422$, p = 0 .025) but not by food logs. When the two groups were examined separately there was no significant correlation between TDEE and intake as assessed by dietary recall in the BED group, but there was a trend that indicated a possible correlation in the control group (n = 13, $r^2 = 0.522$, p = 0.056).

Body composition

There were no differences between groups on baseline measures of fat and lean tissue compartment, follow-up fat and lean tissue compartments or on change in fat, change in lean, according to the DXA scan data. There were also no within group differences in baseline and follow-up on fat or lean tissue compartments. There were no within or between group differences in weight from baseline to follow-up.

Random 24-hour dietary recall data

BED participants had an average of 2.29 binge days during the 6 dietary recalls (median = 2, range = [0,5]). The BED group had a significantly higher caloric intake on days when they had binge eating episodes than on days when they did not (Table 23). Additionally, caloric intake in the BED group on binge days was significantly higher than control average intake. There was no difference between BED non-binge day intake and

control intake (Table 22). There was a trend toward higher average daily intake in the BED group (p=0.053). There was a significant group difference in number of kilocalories consumed per unit of BMI with the BED group consuming 76.2 kcal/BMI unit and the controls consuming 61.0 kcal/BMI unit.

Food log data

BED participants had an average of 7.5 binge days during 14 days of food log entries (median = 7.5, range = [4,11]). The food log data corroborated that BED participants consumed significantly more kcal on binge days than non-binge days (Table 23) and had greater intake on binge days than controls (Table 22). The BED group had similar intake on non-binge days to controls. There were no significant differences in average intake or kcal/BMI unit between groups according to food log data.

Energy expenditure versus reported intake

Daily intake as reported by the 24-hour recall data and the food log data were compared to actual TDEE as assessed by DLW (Table 24). BED participants reported caloric intake that was 80% of TDEE according to dietary recall data and 70% of TDEE according to food log data. Control participants reported caloric intake that was 67% and 72% of TDEE according to dietary recall and food log data, respectively. There were no significant group differences in under-reporting between groups.

Predicted energy requirements versus energy expenditure

There were no between group differences on HBE predicted energy requirements. When comparing predicted energy requirements to actual TDEE there were no group differences. Predicted energy requirements accounted for 76% and 77% of actual TDEE for BED and control groups respectively (Table 24).

Energy and macronutrient intake during an overeating episode: laboratory vs. dietary recall

Table 25 reports descriptive and test statistics for laboratory and dietary recall intake. Total food intake was significantly greater in those with BED than those without according to laboratory (2305.1 vs. 1461.8 kcal; 466.3 vs. 294.4 gm) and dietary recall methodologies (2091.1 vs. 1312.8 kcal; 411.1 vs. 261.6 gm). Compared to overweight/obese controls, those with BED consumed significantly more grams of carbohydrate (laboratory: 294 vs. 71 gm; recall: 251 vs. 151 gm) and grams of fat (laboratory: 96 vs. 63 gm; recall: 99.7 vs. 59 gm) according to both methodologies. There was no significant difference between BED and control participants in protein intake.

The proportion of energy intake from carbohydrate, fat and protein was also examined. There was no significant difference between BED and control groups in the proportion of intake from carbohydrates and fats. Controls consumed a significantly greater proportion of energy intake from protein than those with BED according to the

dietary recall data (15.1% vs. 20.1%). This difference was not significant when measured in the laboratory.

Multiple methods to evaluate accuracy of self-reported food intake data

Paired samples t-tests demonstrated no significant within group differences in total food and macronutrient intake between laboratory and dietary recall methodologies in either BED or control groups (Table 25). One exception was the proportion of total intake from fat, with the BED group reporting to consume more % fat in the dietary recall than was measured by the laboratory (24.1 vs. 19.7% %). Accuracy of reporting was further examined by calculating the ratio of self-reported intake assessed by dietary recall to measured intake in the laboratory (dietary recall / laboratory). The proportion of reported to measured food intake measured in grams was 0.94 (SD = 0.040) in the BED group and 0.89 (SD = 0.36) in the control group. There was no significant difference between groups ($F(1, 28) = 0.147$, $p = 0.704$). Additionally, there was no significant difference between groups when evaluating the dietary recall / laboratory ratio with food intake measured in kcal (0.98, SD = 0.45 vs. 0.90, SD = 0.36; $F(1, 28) = 0.308$, $p = 0.584$). The correlation between laboratory and dietary recall methodologies was calculated to evaluate accuracy of self-reported data. In both BED and obese control groups, significant within group correlations were found between laboratory and recall methods for total food intake measured in kcal (BED: $r = 0.530$, $p=0.05$, CON: $r = 0.805$, $p < 0.001$). Total food intake measured in grams was significantly correlated between methods in the control group ($r = 0.777$, $p < 0.001$), but only a trend towards significance in the BED group ($r = 0.465$, $p = 0.09$). The difference between the two correlation coefficients approached significance (kcal $z = 1.33$, $p = 0.091$, gm $z = 1.36$, $p = 0.086$).

To further explore the accuracy of self-reported intake, we performed a between group comparison of the directional difference and the absolute value of the mean difference for total food and macronutrient intake assessed by laboratory and dietary recall methods (Table 26). Results demonstrated that the BED participants had much greater variability in their self-reported data as can be seen by the standard deviation of the group means and the magnitude of the absolute values of the mean for the majority of the comparisons in Table 26. There was a trend toward the absolute value of the difference being significantly greater in those with BED than obese controls.

Relationship between total food intake and BMI

Examining BED and obese control groups together, BMI and total food intake were not significantly correlated according to laboratory (kcal: $r = 0.192$, $p = 0.292$; gm: $r = 0.199$, $p = 0.274$) or dietary recall methodologies (kcal: $r = 0.214$, $p = 0.265$; gm: $r = 0.206$, $p = 0.284$). In the control group, BMI was significantly correlated with food intake in the laboratory (kcal: $r = 0.541$, $p = 0.025$; gm: $r = 0.562$, $p = 0.019$) and dietary recall data (kcal: $r = 0.540$, $p = 0.038$; gm: $r = 0.543$, $p= 0.036$). In those with BED, BMI and food intake were not significantly correlated according to laboratory (kcal: $r = -0.057$, $p = 0.840$; gm: $r = -0.054$, $p = 0.849$) or dietary recall data (kcal: $r = -0.108$, $p = 0.714$; g: $r = -0.123$, $p = 0.675$).

Patterns of energy consumption throughout the day

Table 27 reports descriptive and test statistics for patterns of energy consumption throughout the day. There were no significant differences between BED and obese control groups in pre-binge or post-binge caloric intake. No significant correlations were found between pre-binge and binge intake or post-binge and binge intake in the BED group. In the obese control group, pre-binge intake was significantly correlated with binge intake ($r = 0.576$, $p = 0.025$) and post-binge intake was marginally significant ($r = 0.505$, $p = 0.055$). Pre-binge and post-binge intake were significantly positively correlated in those with BED ($r = 0.616$, $p = 0.019$), obese controls ($r = 0.564$, $p = 0.028$) and overall ($r = 0.465$, $p = 0.011$).

DISCUSSION

The data do not support the hypothesis of higher energy expenditure in the BED group as there were no statistical differences in TDEE, BMR, or TEF between BED participants and overweight/obese controls. Using the doubly labeled water method in the current study, TDEE was 3214 and 3172 kcal/day in BED and non-BED participants, respectively. To the best of our knowledge, this is the first study to measure TDEE by the DLW method in overweight/obese females with BED. Obesity researchers using DLW to measure TDEE have reported values ranging from 2090 kcal/day in obese females during periods of dietary restraint (35) to 3708 kcal/day in obese females with a mean BMI of 37.4 kg/m2 (36). Examining studies of obese females with a BMI range from 29.6 to 33.0, the reported TDEE ranged from 2452 to 2952 kcal/day (37-41). The high TDEE in our study may be a result of higher BMI in our BED (34.8) and control groups (35.2) that approached that of Platte's participants (37.4) (36). Measurements of BMR in our BED and control groups are consistent with those for obese females reported in the literature ranging from 1502 kcal/day to 1680 kcal/day (37, 39, 40). As stated above, we found no difference between BED and non-BED in the thermic effect of food. Some researchers have demonstrated decreased TEF in obese participants (42), but these findings are controversial as others have found no difference between obese and normal weight individuals. Together these results suggest that there are not significant differences in energy expenditure and metabolic measurements between overweight/obese women with and without BED. Additionally, there was no difference in body composition between groups and no change in body composition over the two weeks of DLW sample collection within either group.

In clinical practice and weight loss programs, many still rely on the Harris-Benedict equation (HBE) to estimate energy requirements. We calculated daily energy expenditure using the HBE and compared it to TDEE measured by the DLW method. The HBE substantially underestimated measured TDEE in this sample by about 23% and 24% in the BED and control groups, respectively. Estimates of predicted energy expenditure calculated using the HBE should be interpreted with caution given this discrepancy. Further research is needed to validate the utility of the Harris-Benedict equation as an estimate of energy expenditure in overweight/obese and eating disordered

individuals. Equations may need to be adjusted for accurate prediction of energy requirements for overweight/obese populations.

A second objective of this study was to assess differences in energy intake between groups. In the current study, the BED group ate significantly more on binge days than on non-binge days and controls. This finding was detected by both laboratory (2305 vs. 1462 kcal) and dietary recall methodologies (2091 vs. 1313 kcal). There was no difference between BED non-binge days and average daily consumption by controls. There was a trend toward the BED group consuming more kilocalories on average than the controls as assessed by 24-hour recall (BED = 2586.9 kcal, SD = 640.1, Control = 2140.0 kcal, SD = 659.1, $F(1,32) = 4.032$, $p = .053$) but not according to food log data. These discrepant results are likely due to the 24-hour recalls being a more accurate account of food intake than the food log entries (24-hour recall estimates were closer to TDEE as determined by DLW). It is also important to note that the BED group consumed significantly more kilocalories per BMI unit than did the control group adding additional support to the finding of higher daily caloric intake in the BED group.

Macronutrient intake data indicated that those with BED eat significantly greater amounts of carbohydrate and fat than obese controls during a laboratory overeating episode. However, there were no differences in the proportion of total energy intake derived from carbohydrates and fat between groups. This suggests that the differences in total carbohydrate and fat intake observed were secondary to increased food intake in BED participants and do not reflect differences in food selection. This is consistent with previous work by our group in which those with BED consumed significantly more total fat than obese controls, but the proportion of energy intake from fat was not significantly different between groups (293). Dietary recall data indicated that control participants consumed a significantly greater proportion of total energy from protein compared to the BED group. However, this difference was not significant according to laboratory measurements, which is the gold standard for measuring dietary intake. Results of our previous study detected no difference in total or proportion of protein intake between groups (293). We suspect this finding represents differences between groups in accuracy of reporting rather than a true difference in macronutrient consumption.

Other research groups have examined macronutrient intake when obese women with and without BED are instructed to overeat in the laboratory. Yanovski found that those with BED consumed significantly more fat (38.9% vs. 33.5%) and less protein (11.4% vs. 15.4%) than obese controls (292). Guss reported that obese women with BED (BMI >28) consumed a significantly greater proportion of energy from fat than normal weight controls (BMI 19-23), but observed no difference between obese women with and without BED (292). In contrast, Goldfein reported no difference in the proportions of macronutrient intake between obese women with and without BED when instructed to overeat in the laboratory (287). Given these findings and those of the present study, it remains unclear whether differences in macronutrient intake exist between obese women with and without BED. The three studies discussed above utilized an identical laboratory paradigm. Direct comparison of these findings to the current study is difficult because macronutrient consumption reflects both food selection and food presentation, which varies by laboratory protocol.

Our results and the literature review above raise important questions. If there are no differences between the BED and control groups metabolically and the BED group

consumes more energy than the control group then over time the BED group should gain more weight. However, we did not find any statistically significant differences in body composition between baseline measures and the two week follow up. If the BED group is actually consuming more energy and the TDEE is not different from controls then it is possible that our method of measuring change in body composition was not sensitive enough to detect increases in body mass over the two week period or we did not have enough power to statistically support such differences between groups. The test-retest differences for duplicate measures on the DXA scanner was <2%, with the ability to detect changes as low as 0.6 kg (SD = 0.023) (44). The change in kg over the two week collection period for the BED group was +0.033 (SD = 1.62) and for the control group was -0.671 (SD = 1.66) which was not statistically different between groups. Additionally, our *post hoc* power to detect a mean difference of this magnitude between groups at an alpha level of 0.05 was 31.2%. Given the body composition change in this sample was within the confidence limits of the DXA scanner and the limited power to detect changes over a small time period, further research is needed to confirm that indeed BED is associated with higher overall caloric intake and weight gain.

A third objective was to determine the accuracy of caloric intake as assessed by dietary recall interview and food log data. This was done by comparing recorded intake with measured energy expenditure (TDEE) obtained from the doubly labeled water method. Since there was no change in body weight or composition as assessed by DXA during the 14 days of doubly water collection, we can assume that energy intake was equal to TDEE. BED participants reported caloric intake that was 80% of TDEE according to dietary recall data and 68% of TDEE according to food log data. Control participants reported caloric intake that was 70% and 73% of TDEE according to dietary recall and food log data, respectively. There were no significant differences between groups by either method. Reports comparing recorded intake in obese individuals to energy expenditure measured by the DLW method suggest that most report intake that is approximately 60% of predicted expenditure (19, 20). Although the expected 60% accuracy was within our 95% confidence region, our estimates on average were greater.

It is possible that our BED and control groups reported intake with greater accuracy as a result of the dietary recall interviews that the women participated in prior to collecting the self-reported food log data. The dietary recall interviews involved a detailed discussion of food intake and portion sizes with expert interviewers. Additionally, participants were required to watch a food record training video immediately prior to the two weeks of food log data collection. These activities may have trained participants in monitoring food consumption, leading to increased accuracy when recording intake in a food log later in the project. This may account for our groups reporting a higher percentage of TDEE than the 60% seen in most studies.

To expand on the findings above, we compared measured food intake in the laboratory to dietary recall estimates of intake to ascertain the accuracy of self-reported data in obese women with and without BED. To the best of our knowledge, this is the first study to compare laboratory and dietary recall measurements of a specific eating episode in adult women with BED. According to dietary recall interviews, BED and obese control groups reported 90% and 98% of measured food intake during an overeating episode in the laboratory, respectively. Furthermore, there were no

differences between groups in the accuracy of self-reported carbohydrate, fat, or protein intake.

When comparing random 24-hour dietary recall data with TDEE assessed using the DLW method in these same participants, we noted that the BED and obese control groups reported daily food intake at 80% and 68% of TDEE, respectively (277) Since accuracy of self-reported data for an isolated laboratory overeating episode was examined in the current study, direct comparisons to the results above cannot be made. Further research is required to determine if these findings can be replicated and, if so, what factors facilitate the observed improvement in reporting. It is possible that the participants in this study may have reported with greater accuracy because the laboratory environment and unique food presentation made the episode more memorable than eating in the natural environment thus resulting in improved recall of intake.

While there is not total consistency throughout the results, the three methods used to examine the accuracy of self-reported food intake taken together suggest that those with BED were less accurate than obese controls. Within group comparisons demonstrated no significant differences between laboratory and dietary recall methods in total food or macronutrient intake in either group. Significant positive correlations between measured and self-reported intake were observed in both groups. However, the correlation coefficients were larger in the control group ($r = 0.805$ vs. 0.530), indicating that they were on average more accurate than those with BED. There was a trend toward a significant difference between these correlation coefficients ($p = 0.09$). We also examined the directional difference and the absolute value of the mean difference between reported and measured intake to examine the direction and magnitude of the inaccuracies in the two groups. Mean differences indicated that both groups under-reported intake, but those with BED did so to a greater extent (215 vs. 160 kcal, $p = 0.021$). The mean of the absolute value of the difference suggests that the BED group tended to be less accurate at reporting their intake overall than controls (779 vs. 438 kcal, $p = 0.061$). The BED group also demonstrated greater variability in reporting as evidenced by standard deviations that were larger than those noted in controls. Overall, these findings are suggestive that those with BED tended to be less accurate with self-reported intake than obese controls.

Our findings suggest that the overweight/obese control participants demonstrated a trend to be more accurate at estimating total energy and macronutrient intake. It is possible that decreased accuracy of dietary recall data in participants with BED may be the result of subjective loss of control and consumption of an extremely large amount of food in a short period of time. Both of these factors may impair awareness of food consumption in BED relative to control participants. Further research is needed to confirm that BED participants are less accurate at reporting food intake than non-BED overweight/obese and to understand the mechanism of impaired accuracy.

Researchers have observed a positive correlation between food intake and BMI when participants with BED were instructed to binge eat in the laboratory (291) In the current study, we sought to confirm that eating in proportion to BMI accounts for the variability in food intake reported in those with BED. Our results indicated that BMI and food intake were significantly correlated in the obese control group, but not in the BED group. These findings are not consistent with those reported by Guss and colleagues (291), who noted a positive correlation between meal size and BMI in the BED group

under binge eating conditions. Significant correlations were not observed in the BED group when they were instructed to eat a normal meal or in obese control participants under binge or normal eating conditions. These results are not consistent with our findings and question the role of BMI in modulating food intake during a single eating episode in those with BED. Methodological differences between these studies make comparison of results difficult and demonstrate the need for further research addressing this issue.

A final aim of this study was to explore the role of reduced caloric intake as a potential precursor to binge eating in BED. To analyze patterns of energy consumption throughout the day, total daily caloric intake assessed by dietary recall was categorized as pre-binge, binge, or post binge intake. There were no significant correlations detected between pre-binge and binge intake or binge and post-binge intake in the BED group. These findings suggest that food intake preceding and following an overeating episode is not associated with food consumption during the overeating episode alone. In contrast, in the obese control group there were positive correlations between all of the comparisons, suggesting that those who eat more before the overeating episode also eat more during and afterwards. Further, significant positive correlations were noted between pre-binge and post-binge food intake in both groups. This observation suggests that those who tended to consume a larger amount of food preceding the overeating episode also tended to eat more following it. Likewise, those who ate less before the overeating episode also ate less following it. Failure to compensate for overeating with reduced dietary intake may contribute to the development of obesity.

Strengths of our study include the multiple methods used to assess energy intake and the use of the gold standard doubly labeled water method to assess energy expenditure. Although the size of the sample is larger than much of the previous work in this area, it is still a limitation of this study. A larger sample size may have clarified the issue of whether there is a significant difference in average daily intake between those with BED and controls. A further limitation is the lack of inclusion of data on physical activity due to participant noncompliance and technical issues with monitors. Although BED is more common in females, another limitation is the lack of inclusion of men. Future studies should include both sexes.

In summary, a major finding of this study is that regardless of the method used to assess intake, both the BED and control groups underestimate their caloric consumption. It is also interesting to note that there is greater disparity in daily caloric intake between the two methods in the BED group than in the controls. However, both groups reported fewer kcal than required to maintain their current weight since the reported intake was less than the TDEE by both methods of assessment. Thus the main positive finding in our study was well summarized by Lichtman et al in their1992 article in which they compared TDEE using DLW with reported intake, “The failure of some obese subjects to lose weight while eating a diet they report as low in calories is due to an energy intake substantially higher than reported and an overestimation of physical activity, not the abnormality in thermogenesis” (35).

The findings of our group and others repeatedly demonstrate increased intake on binge days compared to non-binge days in BED women. This distinguishes those with BED from typical obesity and lends further support to the diagnostic utility of BED and its inclusion in the upcoming DSM-V. Further research will clarify with increasing

precision the quantity, nutrient composition, and food selections that characterize binge eating episodes in BED. Characterizing the eating behaviors associated with BED – both in the laboratory and through self-reported data - will facilitate accurate diagnosis and assessment of treatment response.

ACKNOWLEDGEMENTS

This study was funded by a National Institutes of Health Grant (R01 MH 060199, MO1-RR00400), supported in part by the Minnesota Obesity Center Grant (P30 DK 60456) and National Institute on Drug Abuse (DA-26119). The authors wish to thank Jennifer Hommerding for her assistance with data collection.

TABLES

Table 21: Descriptive statistics and group differences in demographics, energy expenditure and energy intake measures

(*n*=BED/*n*=CON)	BED Mean (SD)	CON Mean (SD)	F	*p*
Age (years) (17/17)	30.8 (7.2)	31.7 (8.5)	0.107	0.745
BMI (kg/m2) (17/17)	34.8 (6.0)	35.2 (6.9)	0.019	0.891
TDEE (15/14)	3213.9 (552.8)	3171.8 (525.3)	0.044	0.835
TEF (15/14)	35.4 (20.0)	29.7 (21.4)	0.534	0.472
BMR (15/14)	1607.7 (246.8)	1628.1 (336.8)	0.035	0.853
24-hour Recall (Kcals) (17/17)	2586.9 (640.1)	2140.0 (659.1)	4.023	0.053
24-hour Recall (Kcals/BMI) (17/17)	76.2 (23.4)	61.0 (14.2)	5.268	0.030
Food Log (Kcals) (14/16)	2234.4 (386.0)	2185.0 (535.4)	0.082	0.777
Food Log (Kcals/BMI) (14/16)	67.5 (17.4)	62.6 (14.0)	0.707	0.408

Note: BED = binge eating disorder, CON = control, *n* = sample size, SD = standard deviation, BMI = body mass index, TDEE = total daily energy expenditure, TEF = thermic effect of food, BMR = basal metabolic rate.

Table 22: Comparison of caloric intake on BED binge days and non-binge days with control data

BED Binge Days	BED Non-Binge Days	Controls	BED Binge vs. Control		BED Non-Binge vs. Control	
Mean (SD) *n*	Mean (SD) *n*	Mean (SD) *n*	F	*p*	F	*p*
24-hour Recall						
3254.5 (520.0) 14	2233.4 (584.0) 17	2140.0 (659.1) 17	26.429	<0.0001	0.191	0.665
Food Log						
2983.0 (432.6) 11	1972.1 (305.0) 14	2185.0 (535.4) 16	16.815	<0.0001	1.721	0.200

Note: BED = binge eating disorder, SD = standard deviation.

Table 23: Comparison of caloric intake on BED binge days with BED non-binge days

BED Binge Days	BED Non-Binge Days	Binge Day vs. Non-Binge Day	
Mean (SD) *n*	Mean (SD) *n*	F	*p*
	24-hour Recall		
3254.5 (520.0) 14	2343.1 (556.6) 14	26.429	<0.0001
	Food Log		
2983.0 (432.6) 11	1972.5 (343.7) 11	16.815	<0.0001

Note: BED = binge eating disorder, SD = standard deviation.

Table 24: Descriptive statistics and group differences in energy expenditure versus reported intake

(*n*=BED/*n*=Control)	BED Mean (SD)	Control Mean (SD)	F	*p*
24-hour Recall/TDEE (15/14)	0.797 (0.23)	0.675 (0.25)	1.885	0.181
Food Log/TDEE (15/13)	0.702 (0.19)	0.725 (0.24)	0.081	0.778
HBE (15/14)	1759.5 (175.4)	1790.7 (257.4)	0.148	0.704
PER (15/14)	2375.3 (236.8)	2417.5 (347.5)	0.148	0.704
PER/TDEE (15/14)	0.757 (0.14)	0.774 (0.12)	0.121	0.731

Note: BED = binge eating disorder, SD = standard deviation, TDEE = total daily energy expenditure, HBE = Harris-Benedict equation, PER = predicted energy requirements based on HBE and light activity.

Table 25: Mean total energy and macronutrient intake during a laboratory over eating episode: Laboratory measurement vs. dietary recall interview

Laboratory test meal			**Dietary recall interview**			**Laboratory vs. Recall**	
BED (n=15)	CON (n=17)		BED (n=14)	CON (n=15)		BED	CON
Mean (SD)	Mean (SD)	F (1,31) (p)	Mean (SD)	Mean (SD)	F (1,28) (p)	t (p)	t (p)
2305.1 (834.0)	1461.8 (641.9)	10.41 (0.003)	2091.1 (1044.1)	1312.8 (847.5)	4.89 (0.036)	0.859 (0.406)	1.240 (0.235)
466.3 (158.2)	293.4 (123.6)	12.02 (0.002)	411.1 (200.0)	261.6 (159.6)	4.98 (0.034)	1.059 (0.309)	1.364 (0.194)
294.1 (97.9)	176.9 (70.9)	15.29 (<0.001)	251.2 (128.1)	150.9 (84.9)	6.26 (0.019)	1.376 (0.192)	1.585 (0.135)
63.7 (6.2)	60.6 (5.0)	2.43 (0.129)	60.8 (8.5)	58.0 (8.3)	0.812 (0.375)	2.015 (0.650)	1.682 (0.115)
1176.2 (391.6)	707.66 (408.9)	15.29 (<0.001)	1004.8 (512.4)	603.6 (339.7)	6.26 (0.019)	1.376 (0.192)	1.585 (0.135)
96.3 (47.5)	62.8 (32.4)	5.57 (0.025)	99.7 (54.6)	59.0 (45.5)	4.78 (0.038)	-0.131 (0.897)	0.563 (0.582)
19.7 (5.7)	21.0 (4.6)	0.471 (0.498)	24.1 (5.1)	21.9 (5.3)	1.32 (0.260)	-2.426 (0.031)	-1.195 (0.252)
867.1(427.4)	565.2 (291.3)	5.57 (0.025)	897.2 (491.2)	531.2 (409.4)	4.79 (0.038)	-0.131 (0.897)	0.563 (0.582)
75.9 (35.1)	53.7 (29.0)	3.84 (0.059)	60.2 (28.8)	51.7 (38.9)	0.438 (0.514)	1.133 (0.278)	0.921 (0.372)
16.5 (5.1)	18.4 (4.1)	1.25 (0.272)	15.1 (4.8)	20.1 (6.3)	5.83 (0.023)	1.117 (0.284)	-0.838 (0.416)
303.5 (140.3)	214.7 (116.1)	3.84 (0.059)	240.7 (115.1)	206.8 (155.8)	0.438 (0.514)	1.133 (0.278)	0.921 (0.372)

Note: BED = binge eating disorder, CON = control, Kcal = kilocalories, CHO = carbohydrate, gm = grams, SD = standard deviation, F = between groups F-test, t = within groups paired sample t-test, Kcal values were estimated from measured macronutrient values in gm using the following standard conversions: 4 kcal/gm CHO, 4 kcal/gm protein, 9 kcal/gm fat.

Table 26: Mean differences of total energy and macronutrient intake between laboratory and dietary recall methodologies

	Directional Difference			**Absolute Value of Mean Difference**		
	BED	CON		BED	CON	
	MD (SD)	MD (SD)	F (1,28) (*p*)	|MD| (SD)	|MD| (SD)	F (1,28) (*p*)
Total kcal	215.7 (939.0)	160.9 (502.4)	6.03 (0.021)	779.1 (527.5)	438.6 (272.1)	3.83 (0.061)
Total grams	54.0 (190.7)	35.4(100.6)	0.11 (0.744)	158.9 (110.9)	87.3 (57.6)	4.87 (0.036)
CHO (g)	42.2 (114.7)	28.0 (68.5)	4.21 (0.050)	97.0 (70.3)	56.8 (45.6)	2.30 (0.141)
Fat (g)	-1.5 (43.2)	3.4 (23.3)	5.00 (0.034)	35.0 (23.4)	19.4 (12.4)	6.99 (0.014)
Protein (g)	13.3 (43.9)	4.02 (16.9)	4.00 (0.055)	30.7 (33.2)	14.1 (9.6)	4.85 (0.036)

Note: BED = binge eating disorder, CON = control, Kcal = kilocalories, g = grams, CHO = carbohydrate, SD = standard deviation, MD = mean difference between laboratory and dietary recall, |MD| = absolute value of mean difference between laboratory and dietary recall.

Table 27: Patterns of energy consumption throughout the day: Pre-binge, binge, and post-binge food intake

Food intake (kcal)	**BED**	**CON**	**BED vs. CON**
	Mean (SD)	Mean (SD)	F (1,28) (*p*)
Pre-binge	1188.9 (449.9)	1038.0 (346.3)	1.032 (0.319)
Binge	2091.1 (1044.1)	1312.8 (847.5)	4.889 (0.036)
Post-binge	182.6 (152.0)	270.7 (338.6)	0.798 (0.380)
Correlations	**BED**	**CON**	**Overall**
	r (*p*)	r (*p*)	r (*p*)
Pre-binge and binge	-0.145 (0.620)	0.576 (0.025)	0.206 (0.284)
Post-binge and binge	0.107 (0.715)	0.505 (0.055)	0.234 (0.222)
Pre and post-binge	0.616 (0.019)	0.564 (0.028)	0.465 (0.011)

Note: BED = binge eating disorder, CON = control, Kcal = kilocalories, SD = standard deviation, r = Pearson's correlation coefficient.

Chapter 6: Binge eating disorder mediates links between symptoms of depression, anxiety, and caloric intake in overweight and obese women

Adapted from: Roseann E. Peterson, Shawn J. Latendresse, Lindsay T. Bartholome, Cortney S. Warren, Nancy C. Raymond. Binge eating disorder mediates links between symptoms of depression, anxiety, and energy intake in overweight and obese women. *Journal of Obesity* [Epub 2012 Apr 12]

ABSTRACT

Despite considerable comorbidity between mood disorders, binge eating disorder (BED) and obesity, the underlying mechanisms remain unresolved. Therefore, the purpose of this study was to examine models by which internalizing behaviors of depression and anxiety influence food intake in overweight/obese women. Thirty-two women (15 BED, 17 controls) participated in a laboratory eating-episode and completed questionnaires assessing symptoms of anxiety and depression. Path analysis was used to test mediation and moderation models to determine the mechanisms by which internalizing-symptoms influenced kilocalorie (kcal) intake. The BED group endorsed significantly more symptoms of depression (10.1 vs. 4.8, p=0.005) and anxiety (8.5 vs. 2.7, p=0.003). Linear regression indicated that BED diagnosis and internalizing-symptoms accounted for 30% of the variance in kcal-intake ($F(3,28)=4.002$, p=0.017). Results from path analysis suggested that BED mediates the influence of internalizing-symptoms on total kcal-intake (empirical $p<0.001$). The associations between internalizing-symptoms and food intake are best described as operating indirectly through a BED diagnosis. This suggests that symptoms of depression and anxiety influence whether one engages in binge eating, which influences kcal-intake. Greater understanding of the mechanisms underlying the associations between mood, binge eating and food intake will facilitate the development of more effective prevention and treatment strategies for both BED and obesity.

INTRODUCTION

Although there is considerable comorbidity between obesity, eating disorders and other major psychiatric disorders, the mechanisms underlying these associations have yet to be resolved. Binge eating disorder (BED), often associated with elevated body weight and mood disorders, is under consideration for inclusion in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V). BED is defined by the DSM-IV as a provisional eating disorder diagnosis characterized by recurrent episodes of binge eating without weight control compensatory behavior and includes: (1) "eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than what most people would eat during a similar period of time and under similar circumstances," and (2) "a sense of lack of control over eating during the episode". In addition, individuals with BED must experience distress about their binge eating and endorse three of the following symptoms: eating more rapidly than normal, eating until uncomfortably full, eating large amounts when not hungry, eating alone because of embarrassment, and feeling disgusted, depressed or guilty about overeating (87).

Although obesity is not a requirement for a BED diagnosis, research indicates that approximately 70% of those meeting criteria for BED are obese (21). While the prevalence of BED in community samples ranges from 2-5%, approximately 30% of obese individuals seeking weight control treatment meet criteria for BED (88, 89). The recurrent overeating that characterizes BED, along with the absence of compensatory behaviors exhibited by those with bulimia nervosa (BN), is most likely responsible for the high frequency of obesity in this group. Laboratory studies have demonstrated that obese BED individuals consume significantly more kilocalories (kcal) during an overeating episode than obese individuals without a BED diagnosis (285, 288, 290, 292, 305, 306, 308-310).

Psychiatric disorders, including depression and anxiety, have been associated with obesity and BED. The lifetime prevalence of major depressive disorder (MDD) and anxiety disorders in the United States is estimated at 17% and 29%, respectively (90). However, within obese populations, reported lifetime prevalence rates are increased to 32.8% for depression and 30.5% for anxiety (20). Additionally, Strine *et al.* found adults with a current or lifetime diagnosis of depression or anxiety were significantly more likely to engage in unhealthy behaviors such as physical inactivity and to be obese (20). Furthermore, research shows obese individuals with comorbid BED have even greater rates of depression and anxiety than obese individuals without BED (21, 295, 311-313). For example, Grilo *et al.* report, in a study of 404 BED patients, that lifetime history estimates were elevated to 52% for mood and 37.1% for anxiety disorders (311).

Despite general acknowledgment of the associations between body weight, BED and comorbid psychiatric disorders, the mechanisms underlying these relationships remain largely unknown. Previously, we have reported that overweight/obese women with BED consume significantly greater kcal-intake during a laboratory eating-episode than weight-matched women without BED (2305 vs. 1462 kcal) (310). To extend this work, we assessed symptoms of depression and anxiety in this sample and sought to examine how internalizing behaviors and BED may be associated with kcal-intake during the laboratory eating-episode. Based on the literature, we hypothesized that participants meeting BED criteria would endorse significantly more symptoms of depression and

anxiety than weight-matched non-BED controls. However, the impact of a BED diagnosis and symptoms of depression and anxiety on kcal-intake was less clear as there are several potential mechanisms responsible for the association. It is possible that increased kcal-intake is the result of BED symptomatology. For instance, those with BED may use binge eating to alleviate or escape symptoms of depression and anxiety. Additionally, in converse, it is possible that BED symptomatology such as distress regarding lack of control over eating specifically elevates internalizing symptoms. For example, depression may increase food intake through increased appetite, a clinical feature of atypical depression subtype. Furthermore, it is possible that having both a BED diagnosis and elevated symptoms of depression and anxiety synergistically influence food intake in a non-additive manner.

A common statistical approach to examining relationships between variables is path analysis, in which alternative models can be applied to evaluate theoretical relationships and determine directionality of effects. We assessed three alternative models, depicted in Figure 13, to determine the mechanism of association that best fit our data. Path analysis was employed to evaluate three potential models: 1) The symptoms of depression and anxiety increase susceptibility to BED, which in turn influences caloric intake (Figure 13a), 2) A BED diagnosis influences symptoms of depression and anxiety, which subsequently influences caloric intake (Figure 13b) and 3) A BED diagnosis and symptoms of depression and anxiety function interdependently in relation to energy intake (Figure 13c).

METHODS

Participants

Participants were recruited by newspaper and online advertisements inviting women at least 50 pounds overweight and between the ages of 18 and 45 to participate in a paid research study. Thirty-two women, including 15 meeting DSM-IV criteria for BED and 17 overweight/obese controls with no history of any binge eating or eating disorder behaviors, participated in the study. These women were recruited as part of a larger study examining food intake and energy expenditure measured via the doubly labeled water method (277, 310).

Group Assignment

Potential participants were interviewed with the Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-I/P) (301) , and the Eating Disorder Examination (EDE), Version 12.0D (303) to determine study eligibility and group assignment. Additionally, a medical history, physical exam and battery of laboratory tests were completed to detect unstable medical conditions, such as diabetes and impaired thyroid function, which would influence eligibility. Participants were excluded from the study if they had any unstable medical or psychiatric conditions, met DSM-IV criteria for substance abuse or dependency within 6 months of participation, or were currently dieting or participating in a weight loss program. Those with any history of BN or

compensatory behaviors were also excluded. Non-BED controls were free of any current or past eating disorder symptoms. The protocol was reviewed and approved by the Institutional Review Board at the University of Minnesota and all participants took part in the informed consent process and signed a consent form. Participants were paid $300 upon completion of the entire study protocol.

Laboratory binge eating episode

This study utilized a protocol our group has previously reported (293, 306, 310). In brief, participants were interviewed by a research dietician regarding their general eating patterns and foods on which they typically snacked or overate. They indicated which items from a standardized list of snack foods appealed to them and could suggest extra foods or recipes. Based on the information gathered during the interview, a tray of binge foods was created for each participant incorporating their personalized snacking preferences. Each participant received 6 to 10 different kinds of food on their snack tray. Food items were presented in excessive quantities (two to three times what they endorsed eating during a binge) to ensure binge size was not limited by quantity of food.

Participants were admitted to the General Clinical Research Center (GCRC) for an overnight stay to participate in several study activities. They were instructed not to consume any food or caloric beverages between 12 and 5 PM. At approximately 5:30 PM they were presented with a multiple item array of foods, including their personalized binge tray and a standard hospital dinner, and were instructed to "Let yourself go and eat as much as you like." They were left alone in a private room to eat for as long as they liked and signaled the nursing staff when they were finished. The GCRC metabolic kitchen staff measured pre and post-prandial quantities of food. Caloric and macronutrient intake for the laboratory eating episodes were calculated using Nutritionist IV (304).

Self report measures of depression and anxiety

During the initial evaluation participants completed the Beck Depression Inventory (BDI) and the Beck Anxiety Inventory (BAI) which are widely used self-report questionnaires consisting of items addressing how one has been feeling in the last week and measures the severity of depression and anxiety symptoms (314, 315). The scales have high internal consistency coefficients (i.e., BDI upwards of 0.80) and validity with other clinical assessments (316, 317). Scores on these indices range from 0 to 63 and correspond to normal (0-9 BDI, 0-7 BAI), mild (10-18 BDI, 8-15 BAI), moderate (19-29 BDI, 16-25 BAI) and severe (30-63 BDI, 26-63 BAI) depression and anxiety.

Analytic strategy and model validation

A set of three, theoretically driven path models (see Figure 13) were tested using Mplus version 5.0 (318). As MacKinnon and colleagues (319) have suggested, the traditional *causal steps approach* (320, 321) may lack the statistical power to detect some meaningful indirect effects. The mediation analyses presented here utilized the *product of coefficients* strategy (319, 322) to evaluate the extent to which a predictor influences

an outcome through some intermediary variable (Figure 13a and b). In doing so, the *indirect effect* is derived by taking a ratio of the product of the path coefficients from (1) the independent variable to the mediator and (2) the mediator to the dependent variable over the normal-theory standard error for that product [i.e., $(\beta 1 * \beta 2)/SE_{(\beta 1 * \beta 2)}$], the results of which are evaluated with respect to the *Z*-distribution. Moderation (Figure 13c) was assessed via the partial path coefficient for a product term (i.e., β/SE_{β}) in the presence of its individual components, and evaluated with respect to a *t*-distribution.

To protect against potential bias introduced by the small size of our sample, evidence of significance was assessed via permutation testing (153). From the original observed data, ten thousand novel datasets were generated via the random reordering of individuals' values on BED and kcal intake. This procedure was performed in R version 2.9.1 (323). Each of the permuted datasets can thus be reanalyzed within Mplus version 5.0 (318), with respect to the three alternative models depicted in Figure 13, and the test statistics from each iteration can be used to generate null distributions for each of the effects being scrutinized. Criteria for significance (i.e., empiric *p*-values) can be calculated using the formula (p+1)/(n+1), where p is the number of null tests that are more significant than the test conducted with the original data, and n is the total number of permutations (i.e., 10,000) on which the analyses are rerun. As a result, we are able to assess whether each of the hypothesized models would achieve significance in a much larger sample (i.e., 320,000), given the characteristics of our observed sample.

RESULTS

Descriptive statistics

Of the thirty-two women participants, 27 were European-American, 3 were African-American (9.4%) and 2 were Asian-American (6.3%). Means and standard deviations for total energy intake, depression and anxiety scores, and potential covariates (i.e., age and BMI) are presented by BED diagnosis on the diagonal in Table 28. ANOVA indicated that there were significant group differences in depression scores (10.1 vs. 4.8, $F(1,30) = 9.308$, $p = 0.005$) and anxiety scores (8.5 vs. 2.7, $F(1,30) = 10.830$, $p = 0.003$) with BED participants having significantly higher mean scores than controls across these indices. No between-group differences were found regarding BMI ($F(1,30) = 3.203$, $p = 0.784$) or age ($F(1,30) = 10.737$, $p = 0.674$). Table 29 reports the prevalence of lifetime clinical depression and anxiety diagnoses by group. The BED group had significantly greater prevalence of mild depression (60 vs. 17.6%, $\chi^2 = 6.10$, $p = 0.014$), mild/moderate anxiety (33.3 vs. 5.9%, $\chi^2 = 3.94$, $p = 0.047$) and anxiety disorders (46.7 vs. 11.8%, $\chi^2 = 4.80$, $p = 0.028$). A detailed examination of food intake and energy expenditure in these participants are reported in additional manuscripts from our group (277, 310)

Pearson's correlation coefficients for bivariate associations between study variables are presented in the off-diagonal cells in Table 28. Within each cell, associations are presented separately for participants diagnosed with BED (top), weight-matched controls (middle) and across the entire sample (bottom). Significant positive correlations were found between kcal intake and depression, kcal intake and anxiety, and depression and anxiety within the full sample. When assessed within groups, no

significant correlations were found except between depression and anxiety scales in the control group. Since neither BED nor internalizing symptoms were associated with age or BMI, these latter variables were not included in the path models described below.

Model fitting

Separate path models were run to test (a) the intermediary role of BED in associations between depression and anxiety symptoms and caloric intake, (b) the intermediary role of depression and anxiety symptoms in associations between BED and caloric intake, and (c) the interactive influences of BED and symptoms of depression and anxiety on caloric intake. In each case, the theoretical model accounted for a significant amount (~30%) of the variance in energy intake. However, examination of the three alternative theoretical models revealed important mechanistic differences in the relationship between BED and symptoms of depression and anxiety as they serve to jointly influence energy intake. Results of the models (Table 30) depicted in Figure 13a and b suggest that while kcal intake is significantly influenced by both depression and anxiety symptoms ($\beta_{total,\ depression} = 0.409$, $p = 0.006$; $\beta_{total,\ anxiety} = 0.399$, $p = 0.003$) and binge eating status ($\beta_{total,\ BED} = -0.508$, $p \leq 0.001$), the effects attributed to symptoms of depression and anxiety operate, in large part, through the influences they have on BED ($\beta_{indirect,\ depression\ via\ BED} = 0.197$, $p = 0.052$; $\beta_{indirect,\ anxiety\ via\ BED} = 0.212$, $p = 0.046$). Note that the sign of the effects reflect coding of 1 for BED and 2 for controls in all analyses. That is, roughly half of the influence of depression (~48%) and anxiety (~53%) on caloric intake is mediated through BED. In contrast, the influence of BED status on caloric intake appears not to be mediated by symptoms of depression or anxiety ($\beta_{indirect,\ BED\ via\ depression} = -0.103$, $p = 0.282$; $\beta_{indirect,\ BED\ via\ anxiety} = -0.096$, $p = 0.329$); rather, those direct effects remained strong ($\beta_{direct,\ BED\ with\ depression} = -0.404$, $p = 0.027$; $\beta_{direct,\ BED\ with\ anxiety} = -0.411$, $p = 0.014$). Results of the model depicted in Figure 13c indicate that BED and symptoms of depression and anxiety do not interdependently influence caloric intake. That is, neither the model including depression, nor the model including anxiety yielded significant partial path coefficients for an interaction between BED and the corresponding depression or anxiety symptoms ($\beta_{BED\ x\ depression} = -0.279$, $p = 0.598$; $\beta_{BED\ x\ anxiety} = -0.268$, $p = 0.609$) after taking into account their combined main effects; in each case, accounting for less than 1% of the total variance.

As described above, post-hoc analyses were conducted with 10,000 permuted datasets to determine whether the results observed with respect to the first theoretical model (i.e., BED mediating the association between symptoms of depression and anxiety and caloric intake) were simply due to chance and/or an artifact of the modest size of the present sample. The null distributions generated from these analyses suggested that the indirect effects of both depression and anxiety through BED were highly significant, as far fewer than 5% of the tests exceeded the *p*-values observed in the original data. In fact, of the 10,000 randomly generated datasets, only seven yielded indirect effects of depression through BED that were more significant than the effect observed in the original data ($p = 0.0008$), with only thirty-eight indirect effects of anxiety on BED exceeding the observed level of significance ($p = 0.0009$).

DISCUSSION

The purpose of this study was to examine models by which internalizing symptoms of depression and anxiety influence food intake in overweight/obese women. Our results indicate that BED women endorse significantly more symptoms of depression and anxiety. Additionally, linear regression indicated that BED diagnosis and internalizing-symptoms accounted for 30% of the variance in kcal-intake. Furthermore, results from path analysis imply that BED mediates the influence of internalizing-symptoms on total kcal-intake, which suggests the associations between internalizing-symptoms and food intake are best described as operating indirectly through a BED diagnosis.

The present study found that overweight/obese women with BED endorsed more symptoms of depression and anxiety than non-BED weight-matched controls. Mean scores for the BDI and the BAI indicated mild depression and anxiety in the BED group but normal levels in the control group. Other studies have found elevated depression and anxiety scores in BED individuals (285, 292, 312, 313, 324, 325). For example, in a study by Fandino *et al.*, depression and anxiety scores were significantly greater in the BED group than the obese control group as assessed by the Symptom Checklist 90 and the BDI (325). The lifetime prevalence of MDD in the BED and control groups was 46.7% and 29.4% respectively. These rates are similar to previous reports in BED (21, 295, 311-313) and non-BED obese groups (20). Lifetime prevalence of anxiety disorders was similar to rates of depression in the BED group (46.7%) but was much lower in the control group (11.8%). It is possible that the lower rates of anxiety disorders in the control group were due to the inclusion of overweight women or was an artifact of the limited sample size.

Laboratory studies have demonstrated that those with BED have greater total food intake than obese controls when instructed to overeat (285, 288, 290, 292, 305, 306, 308-310). Two such studies have reported on both food intake and depression symptoms (285, 292). In a sample of 10 obese BED women and 9 obese controls, Yanovski *et al.*, found that the BED group consumed significantly more kcals (2962 vs 2017) and had significantly greater depression scores as measured by the BDI (18.9 vs 5.4) than controls. Additionally, they observed significant positive correlations between kcal intake and BDI score ($r^2 = 0.41$) and between binge meal energy intake and BDI ($r^2 = 0.28$). Geliebter and colleagues compared consumption of a liquid test-meal for 30 obese BED individuals (18 women) and 55 obese controls (43 women). The BED group consumed significantly more grams (1,032 vs 737) of the liquid test meal and endorsed significantly higher depression scores assessed by the Zung Depression Scale. However, a significant correlation between test meal intake and depression score was not found. The discrepancy could be due to several study design differences, including proportion of BED and control participants, inclusion of men and type of food intake (solid vs. liquid meal).

Furthermore, results from linear regression indicated that BED diagnosis and symptoms of depression and anxiety accounted for a significant amount (~30%) of the variance in caloric intake. However, examination of the three alternative models revealed important mechanistic differences in the relationship between BED, symptoms of depression and anxiety and subsequent energy intake. The model that best fit our data indicated that BED mediated the influence of depression and anxiety symptoms on total

kcal intake (Figure 13a). Specifically, our results suggest that the associations found between symptoms of depression and anxiety and food intake are best described as operating indirectly through a BED diagnosis. That is, symptoms of depression and anxiety influence whether one engages in pathological binge eating, which, in turn, influences caloric intake. Our findings did not support model b (BED predicted symptoms of depression and anxiety which, in turn, influence kcal intake) or model c (a significant interaction between symptoms of depression and anxiety and BED as being predictive of kcal intake).

These results highlight the importance of mood in relation to a BED diagnosis and subsequent caloric intake. Other research has also implicated mood in BED. Telch *et al.* interviewed 60 obese women with BED regarding their definition of binge eating and 33% reported it as eating to regulate negative affect (326). With the advent of Ecological Momentary Assessment procedures (EMA), prospective data on precursors to binge eating in the natural environment have been collected (327-331). A study by Stein et al. found in 33 obese women with BED that negative mood was significantly greater at pre-binge times than at non-binge times and that participants attributed binge eating to mood more frequently than hunger or violation of extreme dietary restraint (abstinence violation) (329). Additionally, a study by Hilbert and Tuschen-Caffier, found that mood preceding a binge eating episode was more negative than mood prior to regular eating or at random assessments in a sample of 20 obese women with BED (330). Furthermore, in a meta-analysis of 36 EMA studies of BED and BN, negative affect was significantly greater preceding binge-eating relative to average affect and affect before regular eating (331). A growing body of literature implicates negative affect as a precursor to binge eating in BED.

The implications of the present study are potentially relevant to the clinical treatment of BED and obesity. Research has indicated that mood and eating disorder diagnoses affect weight loss and other treatment efforts. For example, Pagoto *et al.* reported that both BED and depression were associated with less weight loss and depression was associated with study attrition (332). Furthermore, in BED treatment, depression symptoms have been associated with both attrition from cognitive-behavioral therapy and severity of eating disorder psychopathology (333). The current results suggest that targeting mood may be useful in the treatment of BED and accentuate the importance of considering mood and BED status in weight management.

Among the major strengths of this study were utilizing path analysis to test relationships between BED, symptoms of depression and anxiety and kcal intake as well as using permutation procedures for model validation and statistical support. EMA studies have consistently demonstrated negative affect as a precursor to binge eating (331) in BED. However, these studies have relied on self-report of food intake. Research indicates that obese and BED populations tend to underreport their food intake (276-280, 310, 334, 335). Therefore, a further strength of this work was the inclusion of laboratory measured food intake to avoid inaccuracies often associated with self-report of dietary intake. Potential limitations include limited sample size and age range, exclusion of male participants and use of self-report questionnaires to measure symptoms of depression and anxiety. Future research is warranted to confirm our findings and should seek to compare energy intake and depression and anxiety in both women and men. Greater understanding of the mechanisms underlying the associations of depression and anxiety symptoms,

binge eating and caloric intake will facilitate the development of more effective prevention and treatment strategies for both BED and obesity.

ACKNOWLEDGEMENTS

This study was funded by National Institutes of Health Grants (R01 MH 060199, MO1-RR00400), supported in part by the Minnesota Obesity Center Grant (P30 DK 60456) and National Institute on Drug Abuse (DA-26119). Special thanks to Brion S. Maher, Ph.D., for statistical consultation on permutation procedures and Jennifer Hommerding, Psy.D., and Andrea Loveless, Psy.D., for assistance with data collection.

TABLES AND FIGURES

Figure 13: Theoretical models of the associations between internalizing symptoms, binge eating and caloric intake

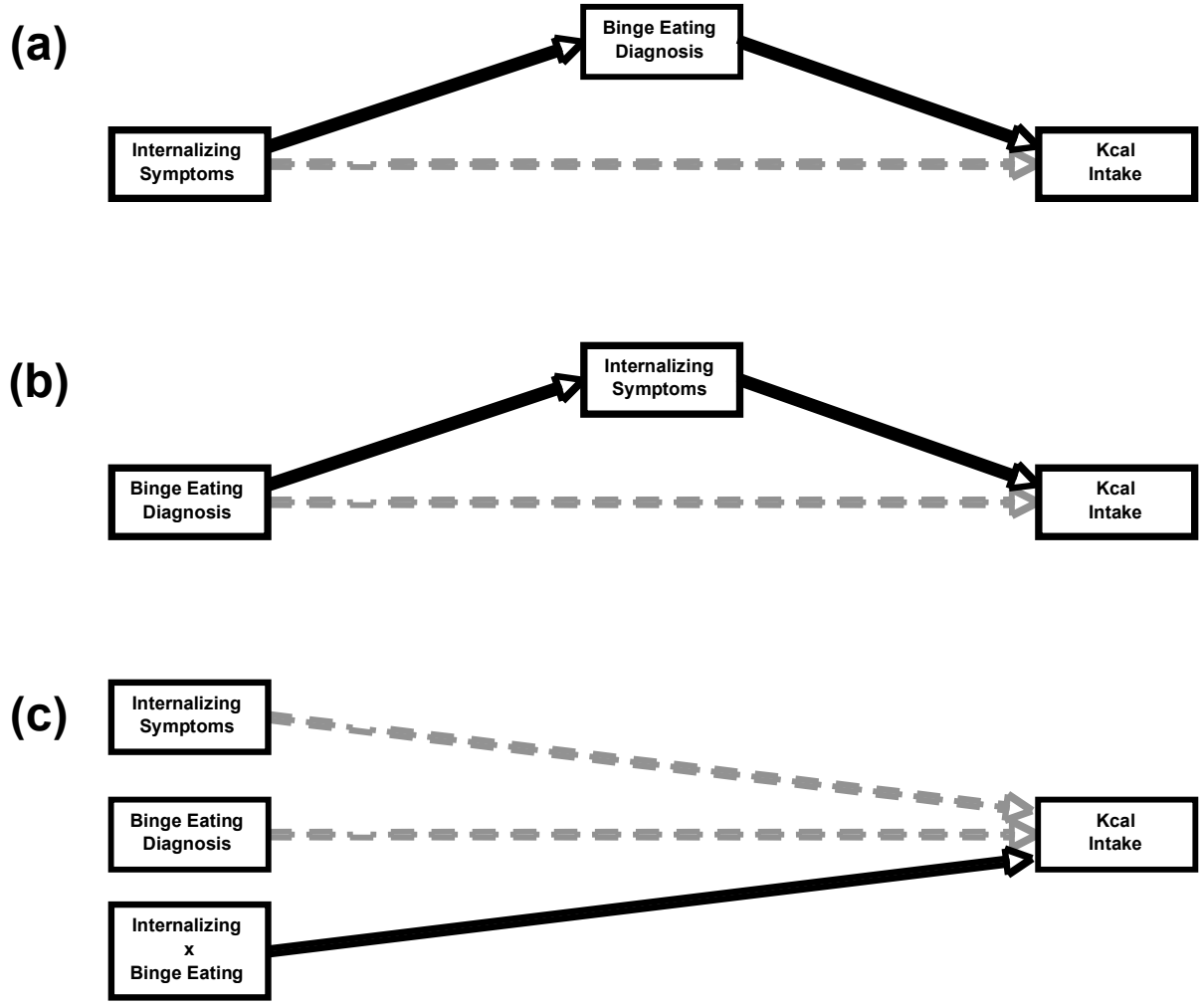


Figure 18. Theoretical models examined: (a) binge eating disorder mediates the associations between internalizing symptoms and kilocalorie intake, (b) internalizing symptoms mediate the association between binge eating disorder and kilocalorie intake, and (c) binge eating disorder interacts with internalizing symptoms in the prediction of kilocalorie intake. Note: Internalizing = symptoms of depression and anxiety.

Table 28: Group means and inter-correlations for study variables

	1	2	3	4	5
1. Age	*30.1 (6.7)* *31.3 (8.5)*				
2. Body Mass Index (kg/m^2)	-0.17 -0.14 -0.15	*34.3 (5.5)* *34.9 (7.2)*			
3. Depression Symptoms	-0.09 0.06 -0.04	-0.33 0.34 0.04	***10.1 (4.8)*** ***4.8 (5.0)***		
4. Anxiety Symptoms	-0.05 0.35 0.04	-0.29 0.24 -0.08	0.34 0.66** 0.57***	***8.5 (6.5)*** ***2.7 (3.1)***	
5. Kilocalorie Intake	-0.10 -0.26 -0.19	-0.06 0.54* 0.02	0.28 0.15 0.41*	0.27 0.00 0.40*	***2305.1 (834.0)*** ***1461.8 (641.9)***

* p < 0.05; ** p < 0.01; *** p < 0.001

Note: Off-diagonal cells depict Pearson's correlation coefficients for participants diagnosed with binge eating disorder (top; n = 15), controls (middle; n = 17), and the overall sample (bottom; n = 32); values on the diagonal reflect means and standard deviations for cases (top) and controls (bottom), with bold-face type indicating group differences ($p < 0.01$) as assessed via F-statistic with 1, 30 degrees of freedom.

Table 29: Lifetime clinical depression and anxiety diagnoses by group

Diagnosis	BED n (%)	Control n (%)	Chi-square	*p*-value
BDI-Mild	9 (60%)	3 (17.6%)	6.10	**0.014**
BAI-Mild/Moderate	5 (33.3%)	1 (5.9%)	3.94	**0.047**
MDD	7 (46.7%)	5 (29.4%)	1.01	0.314
Dep NOS	1 (6.7%)	0 (0%)	1.17	0.279
GAD	1 (6.7%)	0 (0%)	1.17	0.279
Social Phobia	4 (26.7%)	1 (5.9%)	2.61	0.106
Specific Phobia	2 (13.3%)	0 (0%)	2.42	0.120
Panic Disorder	1 (6.7%)	0 (0%)	1.17	0.279
PTSD	1 (6.7%)	0 (0%)	1.17	0.279
Anx NOS	0 (0%)	2 (11.8%)	1.88	0.170
Any Dep Dx	8 (53.3%)	5 (29.4%)	1.89	0.169
Any Anx Dx	7 (46.7%)	2 (11.8%)	4.80	**0.028**
Any Dep/Anx Dx	10 (66.7%)	6 (35.3%)	3.14	0.077

Note: BED = binge eating disorder, Chi-square = Pearson's Chi-square 1 degree of freedom test, BDI-Mild = mild depression as assessed by the Beck Depression Inventory which corresponds to scores 10-18, BAI-Mild/moderate = mild to moderate anxiety as assessed by the Beck Anxiety Inventory which corresponds to scores 8-25, MDD = major depressive disorder, Dep NOS = depressive disorder not otherwise specified, PTSD = post traumatic stress disorder, Anx NOS = anxiety disorder not otherwise specified, Any Dep Dx = any DSM-IV depressive disorder diagnosis, Any Anx Dx = any DSM-IV anxiety disorder diagnosis, Any Dep/Anx Dx = any DSM-IV depressive or anxiety disorder diagnosis, dysthymic disorder and obsessive compulsive disorder were omitted from table because no participants met criteria for these disorders.

Table 30: Standardized effects coefficients, standard errors and corresponding *p*-values for mediation models

Mediator	Total Effect			Direct Effect			Indirect (Mediated) Effect			
	β	*SE*	*p*-value[a]	β	*SE*	*p*-value[a]	β	*SE*	*p*-value[a]	empirical *p*-value[b]
BED	0.409	0.150	0.006	0.212	0.183	0.247	0.197	0.101	0.052	0.0008
BED	0.399	0.136	0.003	0.187	0.181	0.301	0.212	0.106	0.046	0.0009
Depression	-0.508	0.136	< 0.001	-0.404	0.182	0.027	-0.103	0.096	0.282	---
Anxiety	-0.508	0.136	< 0.001	-0.411	0.168	0.014	-0.096	0.099	0.329	---

[a] corresponding to the two-tailed test statistics for models run with sample data.
[b] corresponding to the two-tailed test statistics for a series of analyses with 10,000 permuted datasets.
Note: BED = binge eating disorder, Signs of effects reflect coding of BED status as 1 and control as 2 in all analyses.

Chapter 7: Genetic and environmental associations between body mass index, depression symptoms and impulsivity in a population-based sample of twins: VA30k

Adapted from: On the association of body mass index and depression in a population-based sample of twins. Roseann E. Peterson, B.A., Hermine H. Maes, Ph.D., Lindon J. Eaves, Ph.D., D.Sc., Presentation, June 19, 2009. Behavior Genetics Association. Minneapolis, Minnesota.

ABSTRACT

Obesity and major depressive disorder each represent diseases with complex etiologies which pose a significant burden to public health, affecting 33 and 16 percent of Americans, respectively. Reported heritability estimates are moderate-to-high and studies suggest both positive and negative correlations of these traits. Impulsivity is likely involved in the link between obesity and depression, as it has been associated with each. Despite numerous phenotypic associations between these traits, there has been a lack of reports in the literature investigating genetic and environmental associations between these phenotypes. Therefore, the purpose of this research was to use twin study methodology to investigate if shared genetic and/or environmental liability is potentially responsible for phenotypic associations found between relative body weight, depression symptoms, and impulsivity. Participants were ascertained through the Virginia Twin Registry and a volunteer twin sample solicited through the American Association of Retired Persons (n=14,457 twins, 63.8% female). Female respondents were found to have significantly lower body mas index (BMI) and impulsivity scores (Eysenck Personality Questionnaire), but significantly higher depression symptom scores (Symptoms Checklist) than males. A significant quadratic relationship was found between BMI and depression symptoms, indicating that those with the highest and the lowest BMI were more likely to have greater depression scores. Bivariate twin modeling results did not indicate a significant genetic or environmental correlation between BMI and depression symptoms. However, significant genetic and environmental correlations were found between BMI and impulsivity (rG =0.115, rE=0.046) and a significant genetic correlation between depression and impulsivity (rG=0.075). Trivariate independent pathway twin modeling indicated shared genetic and environmental liability between these traits, although, some sex differences were observed. A common genetic factor accounted for 2-16% of the genetic variance in these traits. For females, an environmental factor common to BMI and impulsivity accounted for 0.5% of the environmental variance in BMI and 62% in impulsivity. For males, an environmental factor common to depression symptoms and impulsivity accounted for 0.5% of the environmental variance in depression symptoms and 56% in impulsivity. Our findings warrant future research in order to confirm these results in additional cohorts as well as to examine how shared genetic risk may impact gene identification efforts.

INTRODUCTION

Obesity and major depressive disorder (MDD) represent serious public health problems and research suggests a prominent sex difference in both, with women appearing to be at increased risk (176, 336). According to the National Center for Health Statistics over 33% of American adults are considered obese while another 33% are overweight (127). Obesity is a general medical condition, defined clinically by a body mass index (BMI) greater than 30 kg/m^2, and is associated with increased risk of numerous medical conditions including cardiovascular disease, insulin-resistance, cancer, and poor quality of life (12, 127). Similarly, depression is a debilitating psychiatric condition that has demonstrated correlations with decreased quality of life, impaired social functioning, eating disorders, substance abuse, and cardiovascular disease (295, 336-338). As reported by the 2006 National Comorbidity Survey Replication, the lifetime history estimates of MDD are 12.7% in men and 21.3% in women (90). However, within obese populations, reported lifetime prevalence rates of depression have been shown to be elevated upwards of 32% (20). Additionally, Strine *et al.* found adults with a current or lifetime diagnosis of depression were significantly more likely to engage in unhealthy behaviors such as physical inactivity and to be obese (20). Cross-sectional studies of BMI and depression have reported positive(93-97), negative (primarily in males) (98, 99) and no association (100-102) between these traits. However, a population based study from the Netherlands found a quadratic (U-shaped) association of BMI and depression indicating those with the lowest and the highest relative body weight were more likely to present with depression. In light of current DSM-IV MDD criteria, which include items related to increase and decrease in appetite, weight and energy expenditure, it is feasible that BMI in underweight and obese individuals may be associated with greater levels of depression (103). Further research is needed to clarify the nature of the association between body weight and depression.

A growing body of research implicates impulsivity in the development and maintenance of obesity. A national study found 17% of the general American population to be impulsive, with odds greater for men and those of younger ages (339). Impulsivity has been considered a multi-dimensional construct consisting of several components: urgency, lack of perseverance, lack of premeditation and sensation seeking (340). Obesity has been associated with dimensions of impulsivity based on both self-report and laboratory-based paradigms (341-348). For example, research conducted using the Iowa Gambling Task has shown obese groups tend to choose immediate rewards, even when future long-term negative consequences are associated with them (342, 345). Furthermore, impulsivity has been associated with dietary disinhibition and may represent a mechanism by which impulsivity may influence body weight via the inability to control what or how much one is eating (349). Dietary disinhibition, a construct from the Three Factor Eating Questionnaire which reflects a responsiveness to food stimuli and eating in response to emotional states (350), has been associated with BMI, obesity, over-eating, decreased healthy food choices, and eating disorders including BED and BN, with less weight loss and with lower levels of physical activity (for a literature review see Bryant *et al.* 2007) (351, 352) .

Impulsivity has been shown to be comorbid with psychiatric disorders. A recent report indicated 83% of those who endorsed impulsivity in a sample of American adults,

also met criteria for lifetime history of at least one psychiatric disorder (339). Research reports positive associations between impulsivity and depression. For example Peluso *et al*., found significantly higher trait impulsivity, as assessed by the Barratt impulsivity scale (BIS), in participants with comorbid bipolar and MDD than controls (353). Additionally, work by our group found a significant positive correlation (r=0.354) between BIS scores and depression symptoms, as measured by the Beck depression inventory, in a sample of obese women with and without binge eating disorder (Peterson *et al.*, in preparation). Furthermore, impulsivity has been shown to be a predictor of future MDD diagnosis (339, 354) and suicidality in depressed persons (355-358).

Research indicates that genetic factors influence individual differences in BMI, depression and impulsivity. Twin, adoption and family studies have consistently shown a significant genetic contribution to body composition with heritability estimates ranging 40 to 70% (34-36). Heritability estimates for depression symptoms have been estimated between 30 and 40% (359) and for MDD between 40 and 50% (360). A large meta-analysis of 11,100 adults indicated impulsivity was moderately heritable, with 31% of the phenotypic variance due to additive genetic effects and 10% to dominance(361). Dietary disinhibition, an impulsivity-associated trait, has been shown to be moderately heritable with 45% of the variance due to additive genetic effects (362).

To date two family studies have examined the genetic and environmental architecture of depression and body composition (363, 364). The first, by Choy *et al.*, was based on 2383 participants from the Netherlands Erasmus Rucphen Family study and did not find a significant genetic correlation between obesity and depression symptoms measured by the Center for Epidemiologic Studies Depression Scale (CES-D) or the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) (363). However, a study by Afari *et al.*, using a sample of 993 female twin pairs from the University of Washington Twin Registry, found significant phenotypic (OR=1.6, 95%CI=[1.2,2.1]) and genetic correlations (12%) between obesity and self-report endorsement of "Has your doctor ever told you that you have depression? (364)." Despite considerable phenotypic associations, there have been no twin and family studies reported in the literature investigating the genetic and environmental associations between BMI and impulsivity or depression and impulsivity. More research is needed to determine if shared genetic and/or environmental liability is responsible for the phenotypic associations found between these traits. Therefore, the purpose of this study was to examine phenotypic associations and the genetic and environmental architecture of BMI, depression symptoms, and impulsivity in a population based sample of twins, the Virginia 30,000 (VA30k).

METHODS

Participants and phenotypes

Ascertainment for the VA30k sample was through two sources, a volunteer twin sample solicited through the American Association of Retired Persons and the Virginia Twin Registry. Participants completed the Health and Lifestyle Questionnaire, which included abbreviated versions of the Symptoms Checklist (SCL-90) and the Eysenck Personality

Questionnaire (EPQ). Depression symptom scores were calculated from 10 questions from the SCL-90 depression sub-scale and impulsivity scores were calculated from 7 items from the EPQ impulsivity subscale. BMI, a standard measure of adiposity, was calculated from self-reported height and current weight. BMI categories were determined from standard clinical groups of underweight (<18.5), normal (18.5-24.9), overweight (25-29.9), obese (30-39.9) and morbidly obese (>40). BMI and depression scores were log transformed and impulsivity scores were arcsine transformed to meet assumptions of normality. Analysis of variance (ANOVA) was used to determine if there were significant differences in study variables by BMI category and age. Tukey's HSD multiple comparison procedure was used to infer in which groups the differences occurred. To explore possible differences in depression symptom profiles between BMI categories ANOVA was applied to individual depression symptom items.

Classical twin methodology

The use of family data allows trait variance to be partitioned into the familial versus residual non-familial sources. In the classical twin design, covariances of MZ and DZ twins are used to estimate the magnitude of genetic and environmental causes of family resemblance (252). This methodology is premised upon monozygotic, or "identical", twins (MZ) sharing all of their genes, while dizygotic, or "fraternal", twins (DZ) sharing half of their genes on average, and MZ and DZ twins sharing trait-relevant environmental experiences to the same extent (equal environment assumption). Following this logic, the correlation between genetic components is modeled as1.0 for MZ twins and 0.5 for DZ twins. Under the assumptions of random mating, no genotype-environment correlation or interaction, and equal environments for MZ and DZ twins, a greater similarity between MZ versus DZ twins is attributed to additive genetic effects (A). Common environmental effects, as defined in biometrical twin modeling, refer to environmental influences that make family members more similar to each other. Therefore, by definition, these influences correlate 1.0 between both MZ and DZ twins. These shared environmental influences (C) will contribute to twin similarity in both MZ and DZ twins and will tend to increase DZ correlations relative to MZ correlations. However, non-additive genetic effects, known as dominance (D), tend to reduce the DZ correlation relative to MZ twins. The correlation of D is modeled as 1.0 between MZ twins and 0.25 for DZ twins. An additional source of variance is the unique environment (E), which includes factors in the environment that are not shared within families as well as random measurement error. Unique environmental influences are uncorrelated between co-twins and have the effect of decreasing the covariance between siblings. Furthermore, the principles of variance decomposition for the univariate case may be extended to estimating the covariance structure between multiple variables.

Model-fit

One approach to partitioning variance is to use structural equation modeling (SEM) and path analysis, which allows for flexible specification of models that include both latent (unobserved) and measured variables (253). In this study, SEM was used to examine the genetic and environmental architecture of BMI, depression symptoms, and impulsivity

for both univariate and multivariate modeling. Model parameters were estimated by full information maximum likelihood using OpenMx(256) in R(210). The goodness-of-fit was assessed by the likelihood-ratio test, which compares minus twice the log likelihood (-2LL) of models. This approximates a χ^2-distribution and may be used for significance testing. Additionally, the relative parsimony of alternative models was assessed by Akaike's Information Criterion (AIC), with smaller values indicating better fit.

Univariate twin modeling and sex-limitation

Univariate models were applied to estimate heritability of individual traits, test the random sample assumption, and determine if there were significant sex differences in the genetic and environmental architecture of the phenotypes. Under the assumption that the twin sample reflects a random sample of the population, there should be no statistical differences on phenotypic mean or variance by twin order or type (zygosity). To test this assumption in the VA30k, phenotypic means and variances were equated by twin order and zygosity to determine if the model fit is significantly worse when compared to the model that estimated them freely. If no differences were found this suggested that the random sample assumptions were met. If significant differences were found, then these could indeed be due to the sample not being random or to some form of social interaction (i.e., sibling cooperation). Furthermore, if there are significant differences in trait variance by gender then it is possible that sex limitation may account for this difference. Two sources of sex limitation are: (1) *quantitative*, also known as scalar sex limitation, defined as sex differences in the magnitude of the genetic or environmental components and (2) *qualitative*, or non-scalar sex limitation, regarded as differences in the actual sets of genes or family environments that influence traits for males and females. For the latter source of sex limitation, addition of opposite sex DZ twins (DZo) to analyses is necessary. In designs of twins reared together, C and D sources of variance cannot be estimated simultaneously. Therefore, ACE and ADE models were tested separately. Along with ACE/ADE models, quantitative and qualitative sex differences were formally tested for BMI, depression symptoms, and impulsivity.

Bivariate twin modeling

To test for genetic and environmental contributions to the covariance between two traits, bivariate Cholesky decomposition was applied. This parameterization allows the phenotypic variance to be partitioned into (1) genetic/environmental components that account for variance in trait one and covariance with trait two and (2) a second genetic/environmental component accounting for the residual variance in the second trait, not accounted for by the first factors. As such, the ordering of the variables determines the interpretation (i.e., how much of the genetic variation in trait two is shared with trait one) [169, 179]. The specification of ACE/ADE models was dependent on best-fit models from univariate modeling. To simplify the full model, A and C/D common and specific factors and E common factors were dropped one-by-one from the model. Specific unique environmental effects were not dropped as these include errors of measurement.

Trivariate twin modeling

To test for shared genetic and environmental liability between BMI, depression symptoms and impulsivity, multivariate Cholesky parameterization and independent pathway (IP) models were fit to the data. Trivariate Cholesky decomposition was used as baseline fit for IP model comparison. As depicted in Figure 26, IP models were specified to partition phenotypic variance into genetic and environmental factors that were shared across all three phenotypes as well as components that were trait specific (243, 253). These models allow for the contributions of the common factors on the measured phenotypes to be different for each of the sources of variance, hence the name 'independent pathways'. IP model fitting began with two common factors for each source of variance, A, D and E, along with specific A, D and E for each variable. To simplify the full model, A and D common and specific factors and E common factors were dropped one-by-one from the model. As noted with previous models, specific unique environmental effects were not dropped as these include errors of measurement.

RESULTS

Phenotypic associations between age, BMI, depression symptoms, and impulsivity

BMI data was available for n=14,457 twins, of whom n=9,227 (63.8%) were female. The mean age was 52.3 and 48.9 years for females and males, respectively. As depicted in Figure 14, females tended to be older than males ($F(1,14357)=119.1$, $p=1.25\times10^{-27}$). Mean BMI was significantly lower for females (23.8 kg/m^2) than for males (25.1 kg/m^2) ($F(1,14455)=310.1$, $p=1.05\times10^{-27}$) and sex accounted for 2% of the phenotypic variance in BMI. Figure 15 displays the distribution of BMI by weight category and sex. Based on a definition of BMI greater than 30 kg/m^2, 12% of the sample was considered obese. A quadratic association was observed between age and BMI, which accounted for 5.5% of the phenotypic variance in BMI (Figure 16).

Depression symptom scores, as assessed by the SCL-90 subscale, indicated that females endorsed significantly higher rates of depression symptoms than males (14.0 females, 13.5 males, $F(1,14118)=306.8$, $p=5.69\times10^{-68}$). Additionally, age was found to be significantly associated with depression symptoms. Specifically, depression scores tended to be greater at younger ages (Figure 17). There was not a significant correlation between BMI and depression scores in females. However, in males a small negative correlation was observed ($r=-0.06$, $p=1.8\times10^{-6}$). As depicted in Figure 19, the depression symptom score was found to have a significant quadratic association with BMI, which accounted for 0.2% of the phenotypic variance. Exploration of depression symptom profiles by BMI category indicated similar endorsement of specific depression items for the underweight and obese groups, except loss of sexual interest, which showed no association with BMI status (Figure 22).

Impulsivity scores, as assessed by the EPQ subscale, indicated that males endorsed significantly higher rates of impulsivity symptoms than females (0.459 females, 0.485 males, $F(1,12670)=26.9$, $p=2.19\times10^{-7}$). In addition, age was found to be significantly associated with impulsivity score, with greater impulsivity observed at younger ages (Figure 18). As depicted in Figure 20, the impulsivity score was found to

have a significant positive association with BMI, which accounted for 0.7% of the phenotypic variance. Additionally, a small but significant correlation was found between depression symptoms and impulsivity in females ($r=0.045$, $p=6.0x10^{-5}$) and males ($r=0.070$, $p=2.2x10^{-6}$). Standardized depression symptom and impulsivity scores are displayed together by BMI category in Figure 21.

Univariate and sex-limitation twin modeling

The phenotypic means and variances of BMI are presented by twin order and zygosity type in Table 31. Means and variances were equated across twin order and across zygosity groups of the same sex without significant loss of model fit, indicating that assumptions regarding random population samples had been met (Table 32). However, means and variances could not be equated between males and females, which was suggestive of possible sex effects. Therefore, sex limitation models were applied in order to test for quantitative and qualitative differences between males and females. Variance component modeling results are displayed in Table 33 (ACE) and Table 34 (ADE). According to the AIC, the best fitting model was an AE model, with the genetic correlation between males and females being estimated. The results indicated that additive genetic effects accounted for 77% of the variance in females and 75% in males. The genetic correlation between males and females was estimated at 0.820 (95% CI = [0.697,0.956]), indicating significant qualitative sex differences although a considerable amount of the additive genetic effects associated with BMI was shared between males and females. These findings suggested that the increased phenotypic variance in females was due, in part, to greater additive genetic variance.

The phenotypic means and variances of depression symptoms, as assessed by the depression subscale of the SCL-90, are presented by twin order and zygosity type in Table 35. Means and variances could be equated across twin order and across zygosity groups of the same sex without significant loss of model fit (Table 36). However, means and variances could not be equated between males and females, which was suggestive of possible sex effects. Therefore, sex limitation models were applied to test for quantitative and qualitative differences between the sexes. Variance component modeling results are displayed in Table 37 (ACE) and Table 38 (ADE). Based on the AIC, the best fitting models were an ACE model in females and an AE model in males, with the genetic correlation between males and females equated to one, which indicates no qualitative sex differences for additive genetic effects. Results indicated additive genetic effects accounted for 28% and 36% of the variance in females and males, respectively, with approximately 8% of the variance in females due to shared environment.

The phenotypic means and variances of impulsivity, as assessed by the subscale of the EPQ, are presented by twin order and zygosity type in Table 39. Means and variances could be equated across twin order and across zygosity groups of the same sex without a significant drop in model fit (Table 40). However, means and variances could not be equated between males and females, suggesting possible sex effects. Therefore, sex limitation models were applied to test for quantitative and qualitative differences between males and females. Variance component modeling results are displayed in Table 41 (ADE). According to the AIC, the best fitting models were an ADE model in females and an AE model in males, with the genetic correlation between males and females

equated to one. The results indicate additive genetic effects accounted for 8% and 32% of the variance in females and males, respectively, with approximately 24% of the variance in females accounted for by dominant genetic effects.

Bivariate twin modeling

Bivariate models were fit to BMI and depression symptoms and results are given in Table 42a. According to the AIC, the best fitting parameterization was model V, which indicated there were no statistically significant shared genetic or environmental liabilities between BMI and depression symptoms. The proportion of variance due to ACE factors is shown in Figure 23.

Variance decomposition models of BMI and impulsivity are displayed in Table 42b. The best fitting model according to AIC was III.b, which indicated that there were statistically significant shared genetic (rG = 0.115, 95%CI = [0.053,0.178]) and environmental correlations (rE = 0.046, 95%CI = [0.011,0.082]) between BMI and impulsivity symptoms. The relative proportion of ADE components are depicted in Figure 24. The proportion of the variance in impulsivity due to genetic effects shared with BMI was 2.8% and 4.4%, in females and males respectively, corresponding to 8.3% and 13% of the total genetic variance. The proportion of the variance in impulsivity due to environmental effects shared with BMI was 1.9% for females (2.8% of total environmental variance) and 1.5% for males (2.3% of total environmental variance).

Results of bivariate model-fitting for depression and impulsivity are displayed in Table 42c. According to AIC, the best fitting model was IV, which indicated a statistically significant shared genetic correlation (rG = 0.075, 95%CI = [0.003,0.151]) between depression and impulsivity symptoms. The variance decomposition is shown in Figure 25. The proportion of the variance in impulsivity due to genetic effects shared with depression symptoms was 1.3% for females and 2.2% for males, which corresponded to 4.0% and 6.5% of the total genetic variance.

Trivariate twin modeling

To test for shared genetic and environmental liability between BMI, depression symptoms and impulsivity, multivariate Cholesky parameterization and independent pathway models were fit to the data. The parameter estimates and fit-statistics are displayed in Table 43 and Table 44. The best fitting model according to the AIC was IP model IV.b and is depicted in Figure 27. The results indicated a significant common genetic factor that loaded on all traits, as well as, genetic effects specific to each phenotype. The common genetic factor exhibited sex differences, in that the factor loaded positively on all traits in females, but in males it loaded positively on impulsivity and BMI but negatively on depression symptoms. In addition, a significant environmental factor was found to load on impulsivity and BMI in females and on impulsivity and depression symptoms in males, further demonstrating sex differences in the nature of these traits.

The proportion of variance in BMI accounted for by ADE components is displayed in Figure 28. In females, 77% of the variance in BMI was due to additive genetic effects, of which 3.8% was due to shared effects with impulsivity and depression

symptoms, 66.7% was specific to BMI, and 29.5% was female specific effects and 23% of the variance was due to environmental components, of which only 0.5% was due to effects shared with impulsivity. In males, 75% of the variance in BMI was due to additive genetic effects, of which 16% was due to shared effects with impulsivity and depression symptoms and 84% was specific to BMI. The remainder of the phenotypic variance was due to a BMI specific environmental component (25%).

The ADE variance decomposition for depression symptoms is shown in Figure 29. In females, 35% of the variance was due to genetic effects, of which 2% was due to an effect shared with BMI and impulsivity, and 65% of the variance was due to environmental factors specific to depression symptoms. In males, the proportion of variance due to genetic factors was 38%, of which, 1.5% was an effect shared with BMI and impulsivity, 36% was accounted for by effects due to dominance, and 62.5% was an effect specific to depression symptoms. The environment accounted for 62% of the variance, of which, 0.5% was due to environmental effects in common with impulsivity.

The proportion of variance in impulsivity symptoms accounted for by ADE components is depicted in Figure 30. In females, 32% of the variance was due to genetic effects, of which 14% was from the common genetic factor, 15% from a specific additive genetic component, and 71% due to specific dominance. The environment accounted for 68% of the variance, of which, 62% was due to the shared factor with BMI and the remainder (38%) was an environmental component specific to impulsivity. In males, 32% of the variance was due to genetic effects, of which, 12% was from the common genetic factor and 88% from an impulsivity specific additive genetic component. The environment accounted for 68% of the variance, of which, 56% was due to the shared factor with depression symptoms and the remainder of the variance from an environmental component specific to impulsivity (44%).

DISCUSSION

The purpose of this research was to examine phenotypic associations between BMI, depression symptoms and impulsivity and to test for shared genetic and environmental liability between these traits in a population-based sample of twins. As expected, our results indicated that women had significantly greater depression symptoms and lower impulsivity than men, and significant positive correlations between BMI and impulsivity, and between depression symptoms and impulsivity were observed. However, rates of obesity in the VA30k sample (12%) were lower than expected given current national estimates (33%). According to national health reports, obesity rates doubled among American adults between 1980 and 2000, which may explain, in part, the lower obesity rate in the VA30k sample, as it was collected during this timeframe.

Reported associations between body weight and depression have been conflicted, with reports of positive, negative and no association between them. Our results indicated a curvilinear relationship between BMI and depression symptoms, signifying that those with the highest and the lowest relative body weight were more likely to endorse depression items. These findings are in agreement with a population-based study from the Netherlands which found a robust U-shaped association between BMI and depression symptoms. It is possible that the mixed findings on the nature of the BMI-depression

relationship may be due, in part, to the assumption that there is a linear association between these traits when indeed it may be curvilinear. Furthermore, examination of depression symptom profiles by weight category indicated similar endorsement of specific depression items for the underweight and obese groups. These findings suggest that there may not be differences in depression profiles by BMI group but rather that those with the highest and the lowest BMIs tend to endorse more symptoms overall. Since both women and those with the highest and lowest BMI's were more likely to have increased depression scores, these groups might be targeted for prevention and intervention efforts.

We applied multivariate twin methods to test for shared genetic and/or environmental liability between these traits. The bivariate twin modeling results did not indicate a significant genetic or environmental correlation between BMI and depression symptoms. Our results are in agreement with a Dutch family study by Choy *et al.* which did not find a significant genetic correlation between BMI and depression symptoms (363). However, Afari *et al.* reported a significant genetic correlation between obesity and self-report endorsement of clinical depression in a sample of female twin from the USA, with this correlation accounting for 12% of the genetic variance (364). It is conceivable that the discrepancies in findings are due in part to different measures of depression (symptoms vs. diagnosis). Further research utilizing genetically informative designs are needed to determine the genetic and environmental structure of comorbidity between body composition and depression and, in particular, whether incorporating depression symptoms versus clinical depression and its subtypes will reveal significant differences in this architecture.

The results from our bivariate analyses on BMI and impulsivity indicated a significant genetic correlation (rG =0.115) between these traits with 8.3% and 13% of the genetic variance in impulsivity due to effects shared with BMI in females and males, respectively. Additionally, a significant environmental correlation (rE=0.046) was also found between these traits indicating ~2.5% of the environmental variance in impulsivity was due to effects shared with BMI. Furthermore, when examining depression symptoms and impulsivity a significant genetic correlation (rG=0.075) was observed, indicating 4% and 6.5% of the genetic variance in impulsivity was due to effects shared with depression symptoms in females and males, respectively. To our knowledge, this is the first twin study to report on shared liability between BMI and impulsivity and between depression symptoms and impulsivity.

The findings from our trivariate twin modeling indicated a significant common genetic factor influencing all three traits. However, we observed significant sex differences, as a positive association was found for all traits for females, but in males this genetic factor was positively associated with BMI and impulsivity but negatively associated with depression symptoms. This suggests that for females, a genetic component exists which is associated with greater impulsivity, BMI and depression symptoms, while in males this genetic component is associated with greater impulsivity and BMI but with decreased depression symptoms. This common genetic factor accounted for different proportions of the genetic variance in each trait as well as some sex differences were observed. The proportion of the genetic variance accounted for by this genetic factor was for BMI 3.8% in females and 16% males; for depression symptoms ~2.5%; and for impulsivity 12-14%. In females, an environmental factor

common to BMI and impulsivity accounted for 0.5% of the environmental variance in BMI and 62% in impulsivity. In males, an environmental factor common to depression symptoms and impulsivity was observed, accounting for 0.5% of the environmental variance in depression symptoms and 56% in impulsivity. Our multivariate twin modeling results suggest that there are shared genetic and environmental factors between BMI, depression symptoms and impulsivity. Further research is warranted to confirm these results in other cohorts as well as to examine how shared genetic and environmental liability may impact gene identification efforts.

A number of extensions to this work should be applied to future research. First, phenotypic associations in this sample indicated a significant quadratic effect of age on BMI as well as significant negative associations with depression symptoms and impulsivity. Future studies should incorporate these effects into modeling, in order to potentially detect differences in genetic and environmental liability by age. Furthermore, BMI and depression symptom scores were also found to have a curvilinear association. There are known limitations of structural equation modeling for the handling of nonlinear relationships. Additional research is needed to determine the effect of nonlinear relationships on variance decomposition methodology and parameter estimates. In addition, since classical twin designs may not model C and D components simultaneously, future models might utilize the extended twin design to determine the effect of each of these sources of variance on the covariance of these traits. Indeed, there are alternative models that may be applied, including models incorporating moderating effects of the environment as well as models of comorbidity (253, 365). For example, longitudinal phenotypic studies have found a reciprocal association between obesity and depression, suggesting that elevated BMI may increase depression and vice versa (91, 92). Therefore, future research should apply models of comorbidity and test direction of causation in a genetically informative sample. To the best of our knowledge, this is the first multivariate twin study to report on the genetic and environmental architecture of BMI, depression symptoms and impulsivity. Our results indicate shared genetic and environmental risk between these traits. Future research is warranted to confirm our findings in additional cohorts and examine how shared genetic risk may impact gene identification efforts.

TABLES AND FIGURES

Figure 14: Percent of sample by age and sex

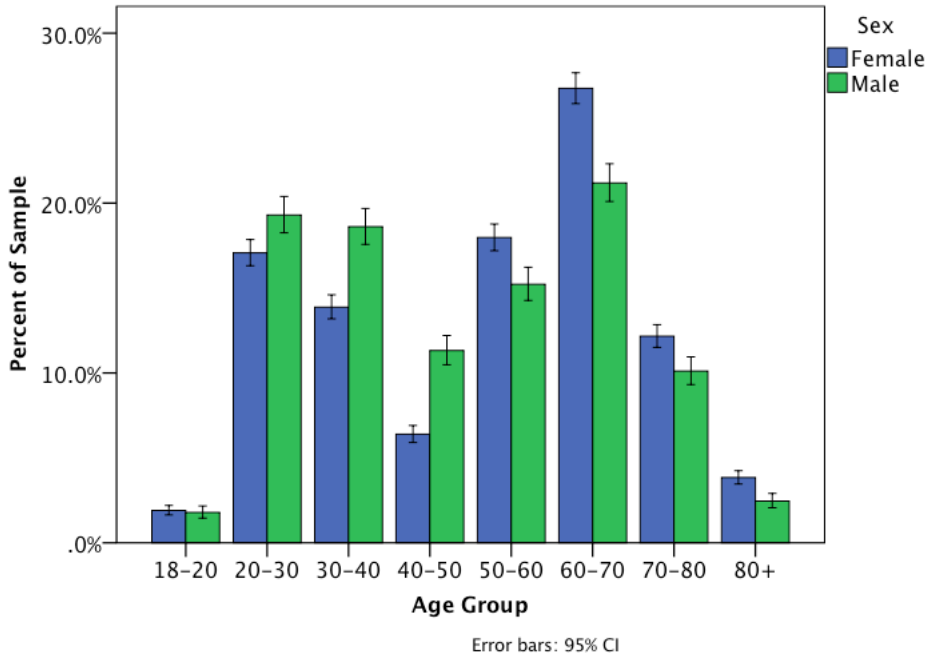


Figure 15: Percent of sample by weight category and sex

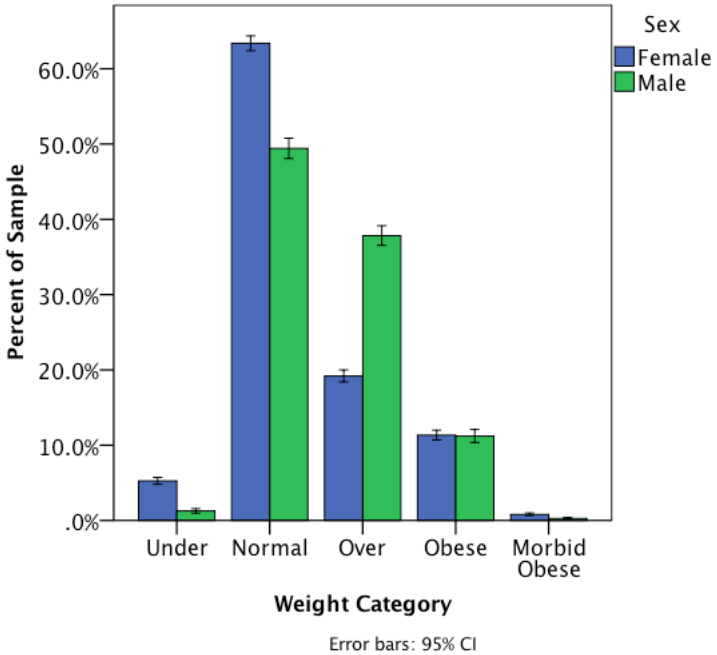


Figure 16: BMI by age and sex

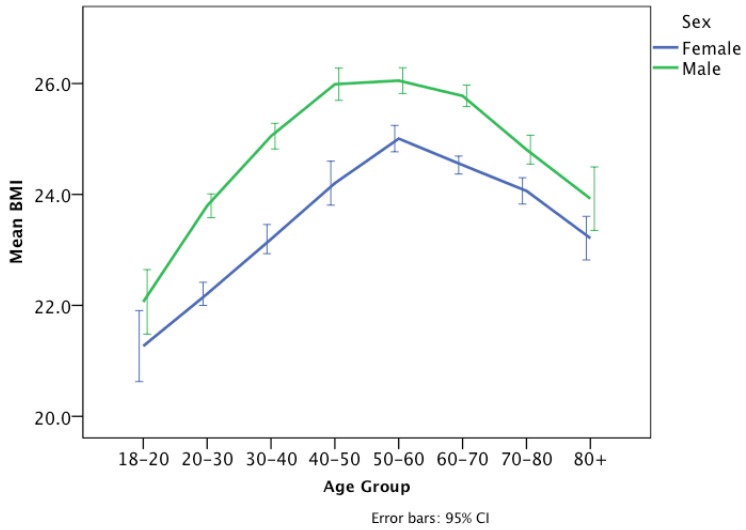


Figure 17: Depression score by age and sex

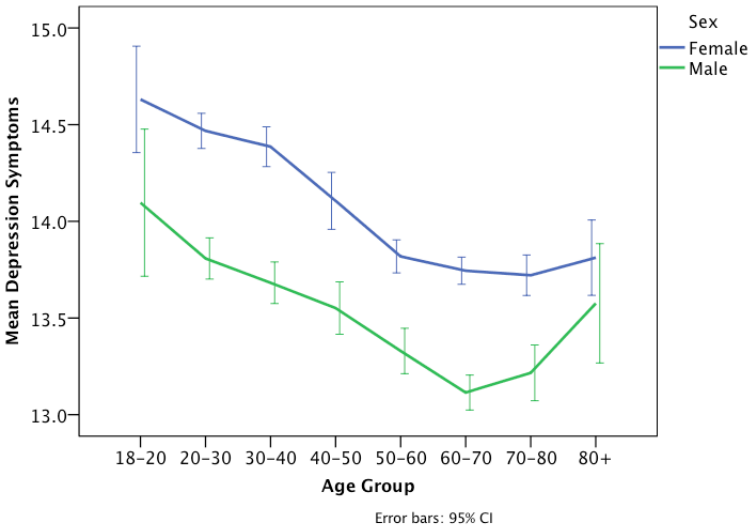


Figure 18: Impulsivity score by age and sex

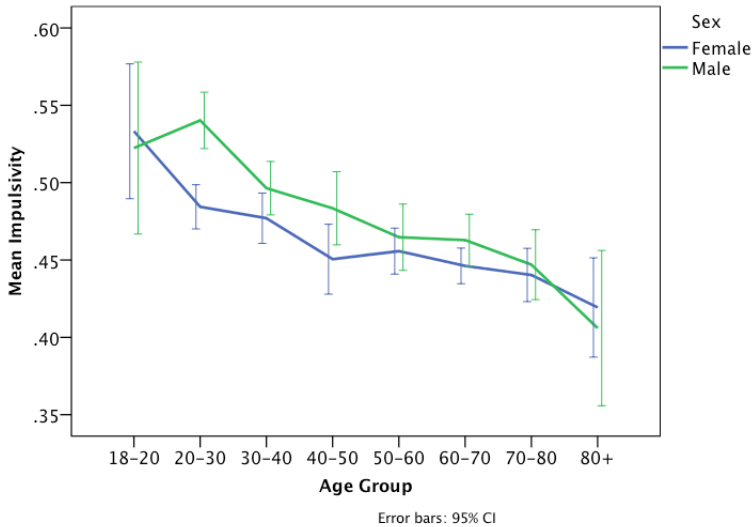


Figure 19: Depression symptoms by weight category and sex

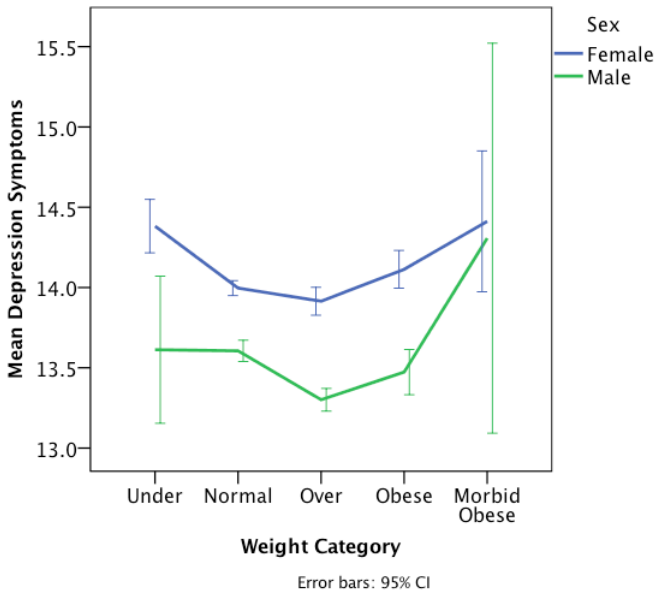


Figure 20: Impulsivity score by weight category and sex

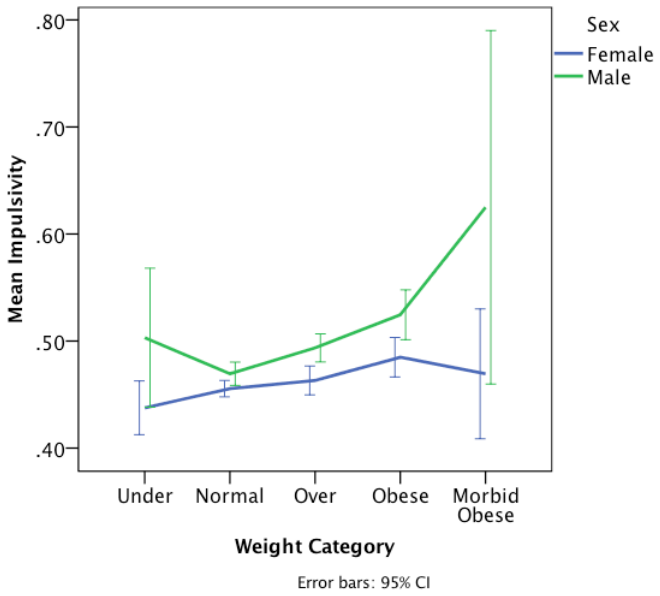


Figure 21: Depression symptoms and impulsivity score by weight category

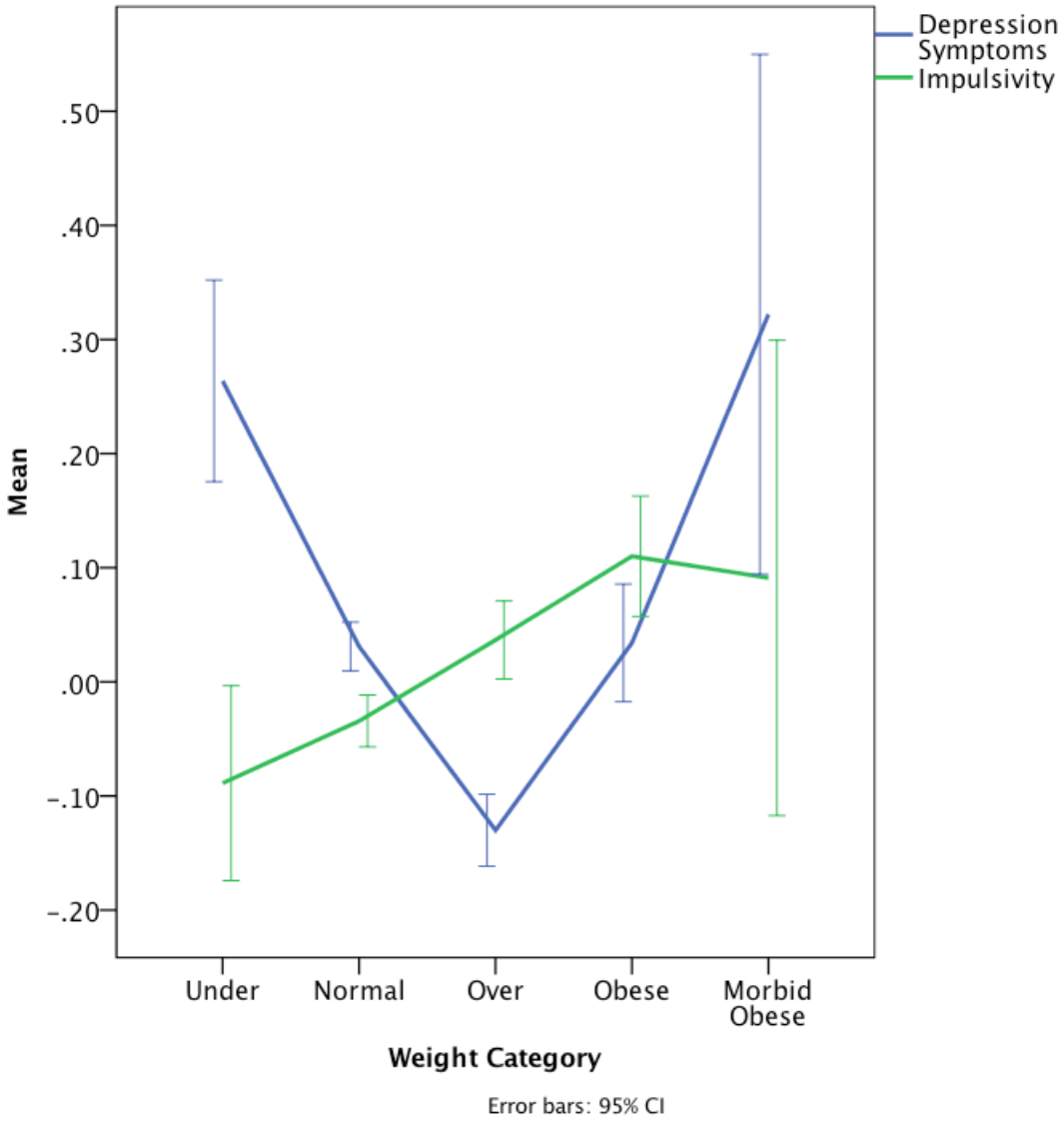


Figure 22: Depression symptom profile by weight category

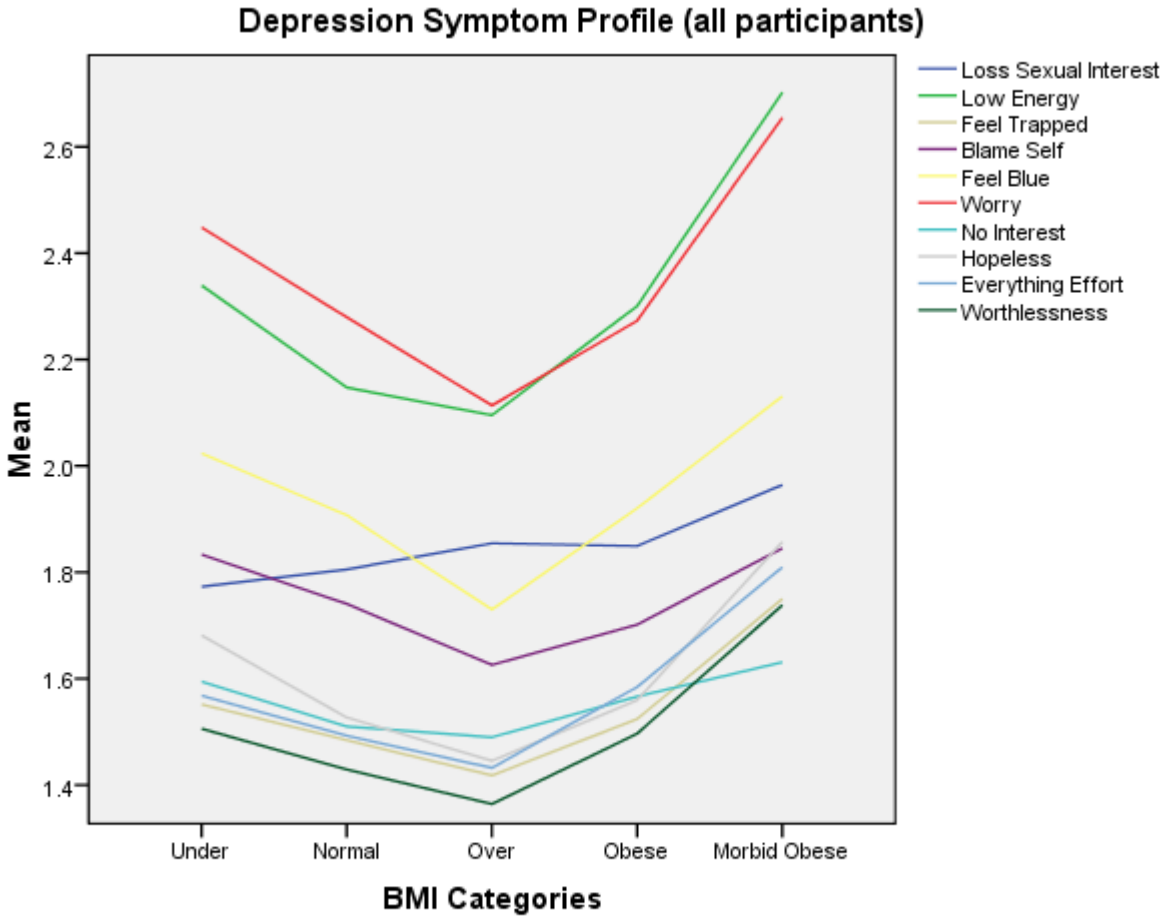


Table 31: Means and variances by twin group for BMI in VA30k

Group (pairs/singletons)	Mean T1 (Variance)	Mean T2 (Variance)	Covariance	Correlation
MZ female (1894/84)	31.51 (3.07)	31.51 (3.00)	2.33	0.772
DZ female (1206/67)	31.61 (3.10)	31.57 (3.05)	1.20	0.392
MZ male (795/18)	32.12 (1.64)	32.08 (1.58)	1.16	0.725
DZ male (590/20)	32.15 (1.80)	32.15 (1.96)	0.70	0.382
DZ opposite sex (1354/43)	32.13 (2.00)	31.38 (2.97)	0.77	0.318

Table 32: Testing model assumptions for BMI VA30k

Model	**EP**	**-2LL**	**Df**	**AIC**	**Diff LL**	**Diff df**	**p-value**
Saturated	25	42190.8	11885	18420.8	-	-	-
Mean order	21	42193.2	11889	18415.2	2.4	4	0.66
Variance order	17	42194.8	11893	18408.8	3.96	8	0.86
Zyg same sex	13	42205.5	11897	18411.5	14.73	12	0.26
Within sex	9	42213.3	11901	18411.3	22.52	16	0.13
Across sex	7	42769.7	11903	18963.7	578.9	18	<0.01

Table 33: Univariate ACE Sex Limitation BMI VA30k

Model	EP	-2LL	df	AIC	Diff LL	Dif df	p-value	FEMALE a A stdA [95%CI]	FEMALE c C stdC [95%CI]	FEMALE e E stdE [95%CI]	MALE a A staA [95%CI]	MALE c C stdC [95%CI]	MALE e E stdE [95%CI]	rg - - [95%CI]
ACE	9	42213.4	11901	18411.4				1.52 2.31 0.763 [0.672,0.785]	0.14 0.02 0.006 [0,0.094]	0.84 0.70 0.231 [0.215,0.248]	1.19 1.41 0.753 [0.628,0.778]	0.0001 0.00 0.000 [0,0.118]	0.68 0.46 0.247 [0.222,0.276]	0.824 - - [0.641,0.999]
AE/ACE	8	42213.4	11902	18409.4	0.02	1	0.89	1.53 2.33 0.769 [0.753,0.785]	- - - -	0.84 0.70 0.231 [0.215,0.248]	1.19 1.41 0.753 [0.628,0.778]	0.0001 0.00 0.000 [0,0.118]	0.68 0.46 0.247 [0.222,0.276]	0.820 - - [0.697,0.962]
ACE/AE	8	42213.4	11902	18409.4	<0.01	1	>0.99	1.52 2.31 0.763 [0.672,0.785]	0.14 0.02 0.006 [0,0.094]	0.84 0.70 0.231 [0.215,0.248]	1.19 1.41 0.753 [0.723,0.778]	- - - -	0.68 0.46 0.247 [0.222,0.276]	0.824 - - [0.697,0.956]
AE	7	42213.4	11903	18407.4	0.02	2	0.99	1.53 2.33 0.769 [0.753,0.785]	- - - -	0.84 0.70 0.231 [0.215,0.248]	1.19 1.41 0.753 [0.723,0.778]	- - - -	0.68 0.46 0.247 [0.222,0.276	0.820 - - [0.697,0.937]
ACE	8	42264.9	11902	18460.9	51.55	1	**<0.01**	1.34 1.79 0.588 [0.531,0.642]	0.74 0.54 0.178 [0.127,0.230]	0.85 0.72 0.235 [0.219,0.252]	0.97 0.95 0.503 [0.421,0.579]	0.67 0.45 0.239 [0.171,0.310]	0.70 0.49 0.258 [0.231,0.290]	0 - - -
ACE	8	42214.5	11902	18410.5	1.16	1	0.28	-1.49 2.23 0.734 [0.661,0.776]	0.33 0.11 0.035 [0,0.105]	-0.84 0.70 0.231 [0.216,0.248]	-1.15 1.33 0.708 [0.611,0.762]	-0.29 0.08 0.043 [0.001,0.135]	0.68 0.47 0.249 [0.223,0.278]	1 - - -
AE/ACE	7	42219.3	11903	18413.3	5.97 4.81	2 1	**0.05** **0.03**	-1.53 2.34 0.769 [0.752,0.784]	- - - -	-0.84 0.70 0.231 [0.216,0.248]	-1.11 1.24 0.656 [0.558,0.758]	-0.42 0.17 0.092 [0,0.184]	0.69 0.48 0.253 [0.226,0.282]	1 - - -
ACE/AE	7	42220.2	11903	18414.2	6.80 5.64	2 1	**0.03** **0.02**	-1.47 2.15 0.704 [0.628,0.777]	0.44 0.19 0.063 [0,0.137]	-0.84 0.71 0.233 [0.217,0.250]	-1.19 1.41 0.750 [0.721,0.776]	- - - -	0.69 0.47 0.250 [0.224,0.279]	1 - - -
AE	6	42222.7	11904	18414.7	9.35 8.19	3 2	**0.02** **0.02**	-1.53 2.34 0.768 [0.751,0.784]	- - - -	-0.84 0.71 0.232 [0.216,0.249]	-1.19 1.41 0.749 [0.720,0.775]	- - - -	0.69 0.47 0.251 [0.225,0.281]	1 - - -

Table 34: Univariate ADE Sex Limitation BMI

Model	EP	-2LL	df	AIC	Diff LL	Dif df	p-value	FEMALE a A stdA [95%CI]	FEMALE d D stdD [95%CI]	FEMALE e E stdE [95%CI]	MALE a A staA [95%CI]	MALE d D stdD [95%CI]	MALE e E stdE [95%CI]	rg - - [95%CI]
ADE	9	42213.3	11901	18411.3	-	-	-	-1.53 2.33 0.769 [0.594,0.785]	0 0 0 [0,0.176]	0.84 0.70 0.231 [0.215,0.247]	-1.17 1.37 0.733 [0.456,0.778]	0.19 0.04 0.020 [0,0.299]	-0.68 0.46 0.247 [0.222,0.275]	0.831 - - [0.697,1.217]
AE/ADE	8	42213.3	11902	18409.3	<0.01	1	>0.99	-1.53 2.33 0.769 [0.753,0.785]	- - - -	0.84 0.70 0.231 [0.215,0.248]	-1.17 1.37 0.733 [0.456,0.778]	0.19 0.04 0.020 [0,0.299]	-0.68 0.46 0.247 [0.222,0.275]	0.831 - - [0.697,1.084]
ADE/AE	8	42213.4	11902	18409.4	0.02	1	0.88	-1.53 2.33 0.769 [0.594,0.785]	0 0 0 [0,0.176]	0.84 0.70 0.231 [0.215,0.247]	-1.19 1.41 0.753 [0.725,0.778]	- - - -	-0.68 0.46 0.247 [0.222,0.276]	0.820 - - [0.697,0.982]
AE	7	42213.4	11903	18407.4	0.02	2	0.99	-1.53 2.33 0.769 [0.753,0.785]	- - - -	0.84 0.70 0.231 [0.215,0.247]	-1.19 1.41 0.753 [0.725,0.778]	- - - -	-0.68 0.46 0.247 [0.222,0.276]	0.820 - - [0.697,0.937]
ADE	8	42301.8	11902	18497.8	88.49	1	**<0.01**	1.02 1.04 0.346 [0.183,0.493]	1.13 1.28 0.426 [0.279,0.589]	0.83 0.69 0.229 [0.214,0.245]	0.42 0.18 0.095 [0,0.322]	1.11 1.22 0.663 [0.435,0.780]	-0.67 0.45 0.242 [0.219,0.269]	0 - - -
ADE	8	42214.4	11902	18410.4	1.03	1	0.31	-1.48 2.19 0.724 [0.570,0.783]	0.37 0.14 0.046 [0,0.200]	0.84 0.70 0.231 [0.215,0.247]	-1.09 1.18 0.633 [0.427,0.762]	-0.48 0.23 0.121 [0,0.329]	0.68 0.46 0.246 [0.221,0.274]	1 - - -
AE/ADE	7	42215.4	11903	18409.4	2.08 1.05	2 1	0.35 0.31	-1.53 2.33 0.769 [0.752,0.785]	- - - -	0.84 0.70 0.231 [0.215,0.248]	-1.02 1.04 0.557 [0.424,0.698]	-0.61 0.37 0.20 [0.054,0.332]	0.68 0.46 0.246 [0.221,0.274]	1 - - -
ADE/AE	7	42218.0	11903	18412.0	4.62 3.59	2 1	0.10 0.06	-1.38 1.90 0.627 [0.501,0.755]	0.66 0.43 0.143 [0.014,0.269]	0.84 0.70 0.230 [0.215,0.247]	-1.19 1.41 0.751 [0.723,0.777]	- - - -	0.68 0.47 0.249 [0.223,0.278]	1 - - -
AE	6	42222.7	11904	18414.7	9.35 8.32	3 2	**0.02** **0.02**	-1.53 2.34 0.769 [0.751,0.784]	- - - -	0.84 0.71 0.232 [0.216,0.249]	-1.19 1.41 0.749 [0.720,0.775]	- - - -	0.69 0.47 0.251 [0.225,0.281]	1 - - -

Table 35: Means and variances of depression symptoms by twin group

Group (pairs/singletons)	Mean T1 (Variance)	Mean T2 (Variance)	Covariance	Correlation
MZ female (1910/67)	0.527 (0.12)	0.521 (0.12)	0.042	0.325
DZ female (1203/69)	0.511 (0.12)	0.538 (0.13)	0.029	0.220
MZ male (805/8)	0.397 (0.09)	0.404 (0.10)	0.033	0.345
DZ male (590/20)	0.431 (0.11)	0.423 (0.10)	0.019	0.159
DZ opposite sex (1362/35)	0.430 (0.11)	0.541 (0.12)	0.019	0.154

Table 36: Testing model assumptions (SCL-90)

Model	**EP**	**-2LL**	**df**	**AIC**	**Diff LL**	**Diff df**	**p-value**
Saturated	25	7546.5	11914	-16281.5	-	-	-
Mean order	21	7552.2	11918	-16283.8	5.75	4	0.22
Variance order	17	7559.6	11922	-16284.5	13.06	8	0.11
Zyg same sex	13	7567.5	11926	-16284.5	20.98	12	0.05
Within sex	9	7577.2	11930	-16282.8	30.69	16	0.01
Across sex	7	7844.8	11932	-16019.2	298.28	18	<0.01

Table 37: Univariate ACE Sex Limitation Depression Symptoms (SCL-90) VA30k

Model	EP	-2LL	df	AIC	Diff LL	Dif df	p-value	FEMALE a A stdA [95%CI]	FEMALE c C stdC [95%CI]	FEMALE e E stdE [95%CI]	MALE a A staA [95%CI]	MALE c C stdC [95%CI]	MALE e E stdE [95%CI]	rg - - [95%CI]
ACE	9	7577.2	11930	-16282.8	-	-	-	0.181 0.033 0.269 [0.145,0.390]	0.105 0.011 0.091 [0,0.197]	-0.278 0.078 0.634 [0.602,0.679]	-0.195 0.038 <0.001 [0.177,0.417]	0.009 <0.001 0.0008 [0,0.153]	-0.258 0.067 0.638 [0.583,0.700]	-0.999 - - [-0.999,-0.119]
AE/ACE	8	7579.9	11931	-16282.1	2.67	1	0.10	0.212 0.045 0.371 [0.334,0.406]	- - - -	-0.276 0.076 0.629 [0.594,0.666]	-0.195 0.038 0.362 [0.206,0.417]	<0.001 <0.001 0 [0,0.134]	-0.258 0. 0.638 [0.583,0.699]	-0.898 - - [-0.999,-0.615]
ACE/AE	8	7577.3	11931	-16284.7	0.07	1	0.79	0.184 0.034 0.279 [0.180,0.391]	0.101 0.010 0.083 [0,0.170]	-0.278 0.077 0.638 [0.602,0.674]	-0.195 0.038 0.364 [0.306,0.418]	- - - -	-0.258 0.067 0.636 [0.582,0.694]	-0.999 - - [-0.999,0.686]
AE	7	7579.9	11932	-16284.1	2.67	2	0.26	0.212 0.045 0.371 [0.334,0.406]	- - - -	-0.276 0.076 0.629 [0.594,0.666]	-0.195 0.038 0.362 [0.303,0.417]	- - - -	-0.258 0.067 0.638 [0.583,0.697]	-0.898 - - [-0.999,-0.615]
ACE	8	7582.1	11931	-16279.9	4.88	1	**0.03**	-0.159 0.025 0.207 [0.104,0.296]	-0.133 0.018 0.146 [0.075,0.232]	-0.280 0.078 0.646 [0.609,0.685]	-0.149 0.022 0.213 [0.084,0.314]	-0.117 0.014 0.130 [0.060,0.229]	-0.262 0.069 0.657 [0.600,0.719]	0 - - -
ACE	8	7577.2	11931	-16284.8	<0.01	1	>0.99	-0.181 0.033 0.269 [0.145,0.387]	-0.105 0.011 0.091 [0,0.197]	-0.278 0.078 0.639 [0.602,0.679]	-0.195 0.038 0.361 [0.255,0.417]	-0.009 <0.001 0.0008 [0,0.087]	-0.258 0.067 0.638 [0.583,0.697]	1 - - -
AE/ACE	7	7580.2	11932	-16283.8	3.00 3.00	2 1	0.22 0.08	-0.212 0.045 0.370 [0.333,0.404]	- - - -	-0.277 0.076 0.631 [0.596,0.667]	-0.186 0.035 0.330 [0.200,0.410]	-0.051 0.003 0.025 [0,0.139]	-0.260 0.068 0.645 [0.590,0.702]	1 - - -
AE/AE	7	7577.3	11932	-16286.7	0.07 0.07	2 1	0.97 0.80	-0.184 0.034 0.279 [0.180,0.384]	-0.101 0.010 0.084 [0,0.170]	-0.278 0.077 0.638 [0.602,0.674]	-0.195 0.038 0.364 [0.306,0.418]	- - - -	-0.258 0.067 0.637 [0.582,0.694]	1 - - -
A	6	7580.4	11933	-16285.7	3.16 3.16	3 2	0.37 0.21	-0.211 0.045 0.369 [0.333,0.404]	- - - -	-0.277 0.077 0.631 [0.597,0.667]	-0.193 0.037 0.357 [0.300,0.411]	- - - -	-0.260 0.067 0.643 [0.589,0.700]	1 - - -

Table 38: Univariate ADE Sex Limitation Depression Symptoms (SCL-90) VA30k

Model	EP	-2LL	Df	AIC	Diff LL	Dif df	p-value	FEMALE a A stdA [95%CI]	FEMALE d D stdD [95%CI]	FEMALE e E stdE [95%CI]	MALE a A staA [95%CI]	MALE d D stdD [95%CI]	MALE E E stdE [95%CI]	rg - - [95%CI]
ADE	9	7579.85	11930	-16280.1	-	-	-	0.212 0.045 0.371 [0.268,0.406]	0.000 0.000 0.000 [0,0.105]	-0.276 0.076 0.629 [0.594,0.666]	-0.191 0.036 0.347 [0.147,0.417]	0.041 0.002 0.016 [0,0.234]	0.258 0.067 0.637 [0.579,0.697]	-0.917 - - [-0.999,-0.615]
AE/ADE	8	7579.85	11931	-16282.1	0	1	>0.99	0. 212 0.045 0. 371 [0.334,0.406]	- - - -	-0.276 0. 076 0. 629 [0.594,0.666]	-0.191 0. 036 0. 347 [0.163,0.417]	0. 041 0. 002 0.016 [0,0.218]	0. 258 0. 067 0. 637 [0.579,0.697]	-0.917 - - [-0.999,-0.615]
ADE/AE	8	7579.86	11931	-16282.1	0.01	1	0.92	0. 212 0. 045 0. 371 [0.270,0.406]	0.000 0.000 0.000 [0,0.103]	-0.276 0. 076 0. 629 [0.594,0.666]	-0.195 0.038 0. 362 [0.303,0.417]	- - - -	0. 258 0. 067 0. 638 [0.583,0.697]	-0.898 - - [-0.999,-0.615]
AE	7	7579.86	11932	-16284.1	0.01	2	>0.99	0. 212 0. 045 0. 371 [0.334,0.406]	- - - -	-0. 276 0. 076 0. 629 [0.594,0.666]	-0. 195 0. 038 0. 362 [0.303,0.417]	- - - -	0. 258 0. 067 0. 638 [0.583,0.697]	-0. 898 - - [-0.999,-0.615]
ADE	8	7606.3	11931	-16255.6	26.49	1	**<0.001**	0.159 0.025 0.209 [0.058,0.330]	0.144 0.021 0.17 [0.050,0.325]	-0.274 0.075 0.621 [0.586,0.657]	-0.028 0.001 0.007 [0,0.239]	0.198 0.039 0.377 [0.139,0.438]	-0.254 0.064 0.616 [0.562,0.674]	0 - - -
ADE	8	7579.94	11931	-16282.1	0.09	1	0.77	-0.212 0.045 0.37 [0.266,0.405]	-0.004 0.000 0.000 [0,0.107]	0.276 0.076 0.63 [0.595,0.666]	-0.179 0.032 0.308 [0.145,0.408]	0.077 0.006 0.057 [0,0.236]	0.258 0.067 0.635 [0.578,0.697]	1.0 - - -
AE/ADE	7	7579.94	11932	-16284.1	0.09 <0.01	2 1	0.96 0.95	-0.212 0.045 0.37 [0.334,0.405]	- - - -	0.276 0.076 0.63 [0.595,0.666]	-0.179 0. 032 0. 306 [0.161,0.408]	0.078 0. 006 0. 059 [0,0.220]	0.258 0. 067 0. 635 [0.578,0.697]	1.0 - - -
ADE/AE	7	7580.35	11932	-16283.6	0.5 0.42	2 1	0.78 0.52	-0.211 0.045 0.369 [0.261,0.403]	0.000 0.000 0.000 [0,0.113]	0. 277 0. 077 0. 631 [0.597,0.667]	-0.193 0. 037 0. 357 [0.300,0.411]	- - - -	0. 260 0. 067 0. 643 [0.589,0.700]	1.0 - - -
AE	6	7580.35	11933	-16285.6	0.5 0.42	3 2	0.92 0.81	-0.211 0.045 0.369 [0.333,0.403]	- - - -	0. 277 0. 077 0. 631 [0.597,0.667]	-0.193 0. 037 0. 357 [0.300,0.411]	- - - -	0. 260 0. 067 0. 643 [0.589,0.700]	1.0 - - -

Table 39: Means and variances of impulsivity (EPQ) by twin group VA30k

Group (pairs/singletons)	**Mean T1** (Variance)	**Mean T2** (Variance)	**Covariance**	**Correlation**
MZ female (1929/49)	0.373 (0.061)	0.379 (0.057)	0.019	0.323
DZ female (1226/45)	0.380 (0.064)	0.373 (0.058)	0.007	0.113
MZ male (803/7)	0.412 (0.059)	0.400 (0.056)	0.018	0.313
DZ male (588/19)	0.426 (0.057)	0.410 (0.063)	0.012	0.189
DZ opposite sex (1380/17)	0.396 (0.059)	0.375 (0.059)	0.004	0.072

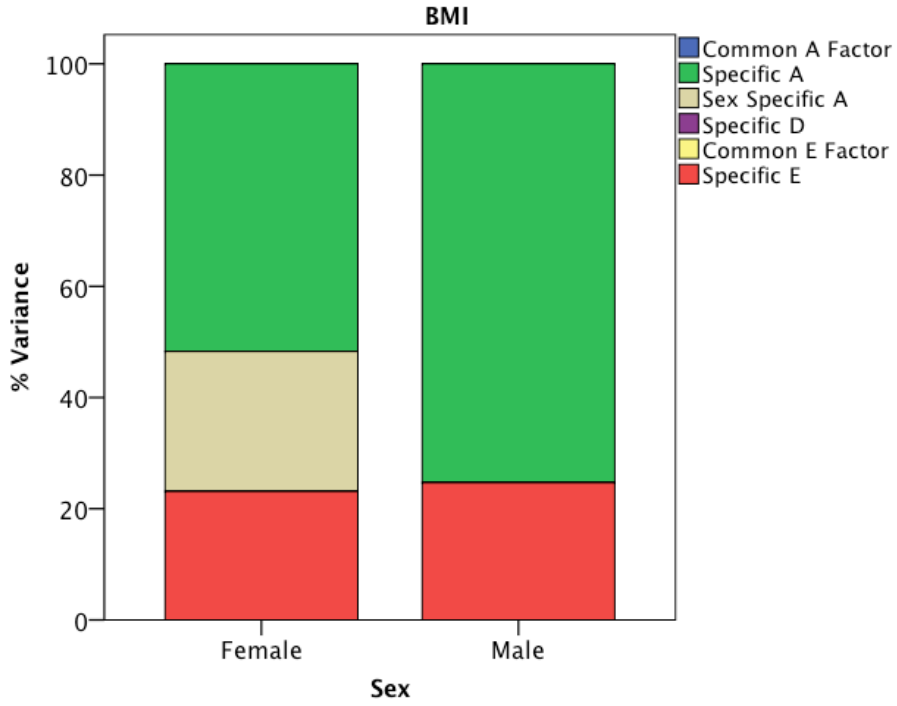
Table 40: Testing model assumptions for impulsivity (EPQ)

Model	**EP**	**-2LL**	**df**	**AIC**	**Diff LL**	**Diff df**	**p-value**
Saturated	25	-193.3	11964	-24121.3	-	-	-
Mean order	21	-188.9	11968	-24124.9	4.42	4	0.35
Variance order	17	-182.4	11972	-24126.4	10.90	8	0.21
Zyg same sex	13	-180.2	11976	-24132.2	13.09	12	0.36
Within sex	9	-176.8	11980	-24136.8	16.58	16	0.41
Across sex	7	-141.7	11982	-24105.8	51.56	18	<0.001

Table 41: ADE models impulsivity (EPQ) VA30k

Model	EP	-2LL	df	AIC	Diff LL	Dif df	p-value	FEMALE a A stdA [95%CI]	FEMALE d D stdD [95%CI]	FEMALE e E stdE [95%CI]	MALE a A staA [95%CI]	MALE d D stdD [95%CI]	MALE e E stdE [95%CI]	rg - - [95%CI]
ADE	9	-175.9	11980	-24135.9	-	-	-	0.084 0.007 0.119 [0,0.325]	0.109 0.012 0.200 [0,0.341]	-0.202 0.041 0.681 [0.643,0.721]	-0.137 0.019 0.320 [0.118,0.375]	<0.0001 <0.0001 0 [0,0.170]	0.199 0.040 0.680 [0.625,0.738]	-0.756 - - [-0.999,0.999]
AE/ADE	8	-172.9	11981	-24134.9	3.00	1	0.08	0.135 0.018 0.307 [0.269,0.344]	- - - -	-0.204 0.041 0.693 [0.656,0.731]	-0.137 0.019 0.320 [0.118,0.375]	<0.0001 <0.0001 0 [0,0.210]	0.199 0.040 0.680 [0.625,0.738]	-0.470 - - [-0.920,-0.131]
ADE/AE	8	-175.9	11981	-24137.9	<0.01	1	>0.99	0.084 0.007 0.119 [0.010,0.325]	0.109 0.012 0.200 [0,0.326]	-0.202 0.041 0.681 [0.643,0.721]	-0.137 0.019 0.320 [0.262,0.375]	- - - -	0.199 0.040 0.680 [0.625,0.738]	-0.756 - - [-0.999,-0.188]
AE	7	-172.9	11982	-24136.9	3.00	2	0.22	0.135 0.018 0.307 [0.269,0.344]	- - - -	-0.204 0.041 0.693 [0.656,0.731]	-0.137 0.019 0.320 [0.262,0.375]	- - - -	0.199 0.040 0.680 [0.625,0.738]	-0.470 - - [-0.815,-0.131]
ADE	8	-172.6	11981	-24134.6	3.25	1	0.07	0.065 0.004 0.071 [0,0.259]	0.122 0.015 0.251 [0.056,0.359]	0.201 0.041 0.678 [0.640,0.718]	-0.115 0.013 0.225 [0.015,0.348]	0.077 0.006 0.102 [0,0.318]	-0.199 0.039 0.674 [0.618,0.732]	0 - - -
ADE	8	-175.8	11981	-24137.8	0.12	1	0.72	0.073 0.005 0.089 [0,0.258]	0.117 0.014 0.231 [0.055,0.342]	0.202 0.041 0.680 [0.642,0.720]	0.136 0.019 0.317 [0.155,0.374]	-0.014 0.0002 0.003 [0,0.173]	0.200 0.040 0.680 [0.625,0.738]	1 - - -
AE/ADE	7	-168.4	11982	-24132.4	7.48 7.36	2 1	**0.02** **0.01**	0.135 0.018 0.305 [0.266,0.343]	- - - -	0.204 0.042 0.695 [0.657,0.734]	0.086 0.007 0.125 [0.021,0.293]	-0.107 0.012 0.197 [0.015,0.324]	0.199 0.040 0.678 [0.621,0.739]	1 - - -
ADE/AE	7	-175.7	11982	-24139.7	0.17 0.05	2 1	0.92 0.83	0.068 0.005 0.077 [0.009,0.195]	0.121 0.015 0.243 [0.116,0.327]	0.202 0.041 0.680 [0.642,0.719]	0.137 0.019 0.320 [0.262,0.375]	- - - -	0.200 0.040 0.680 [0.625,0.739]	1 - - -
ADE/AE	6	-163.9	11983	-24129.9	11.94 11.81	3 2	**0.01** **<0.01**	0.133 0.018 0.297 [0.259,0.335]	- - - -	0.205 0.042 0.703 [0.665,0.741]	0.132 0.017 0.298 [0.239,0.354]	- - - -	0.203 0.041 0.702 [0.646,0.761]	1 - - -

Figure 23: Proportion of variance in BMI and depression symptoms due to ACE components (Bivariate)



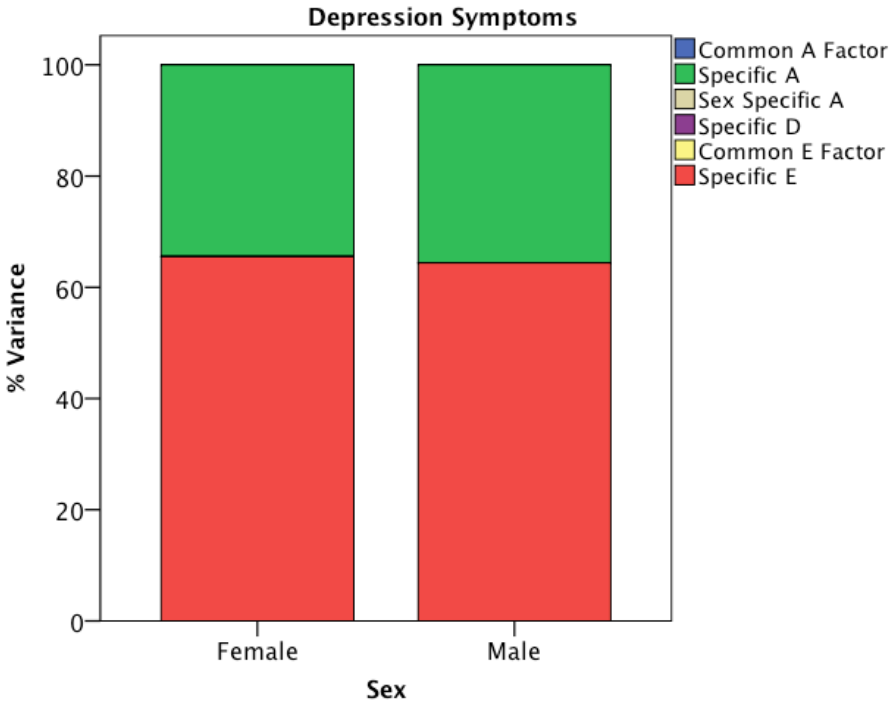
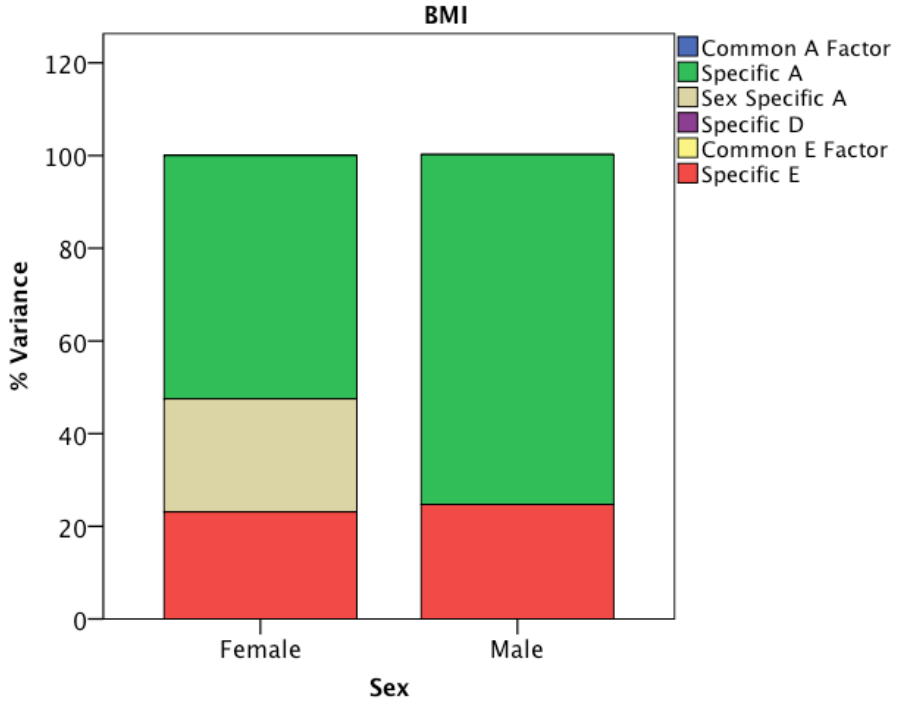


Figure 24: Proportion of variance in BMI and impulsivity symptoms due to ADE components (Bivariate)



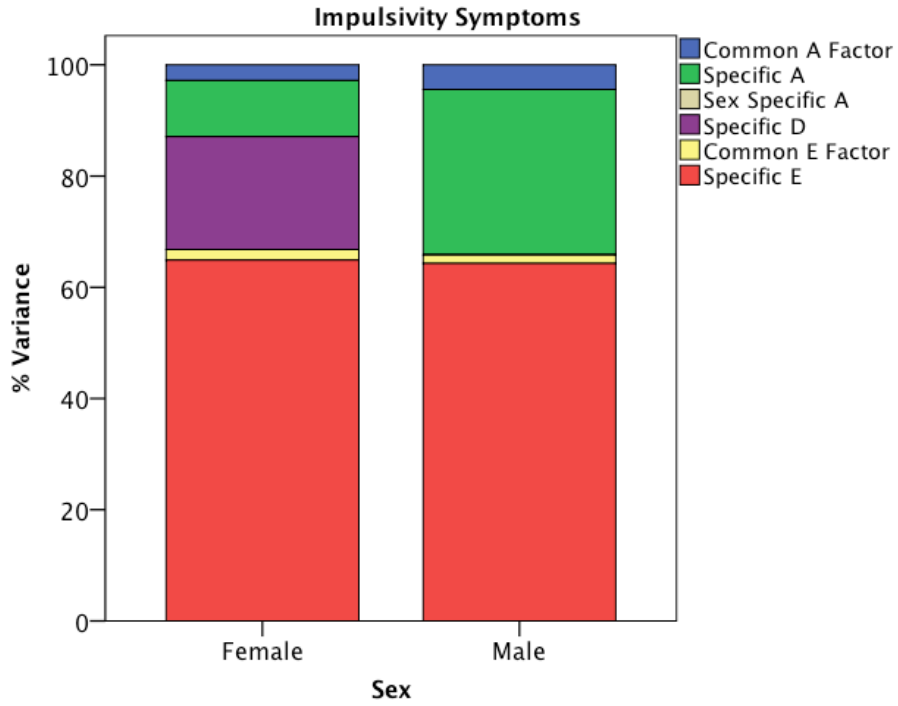
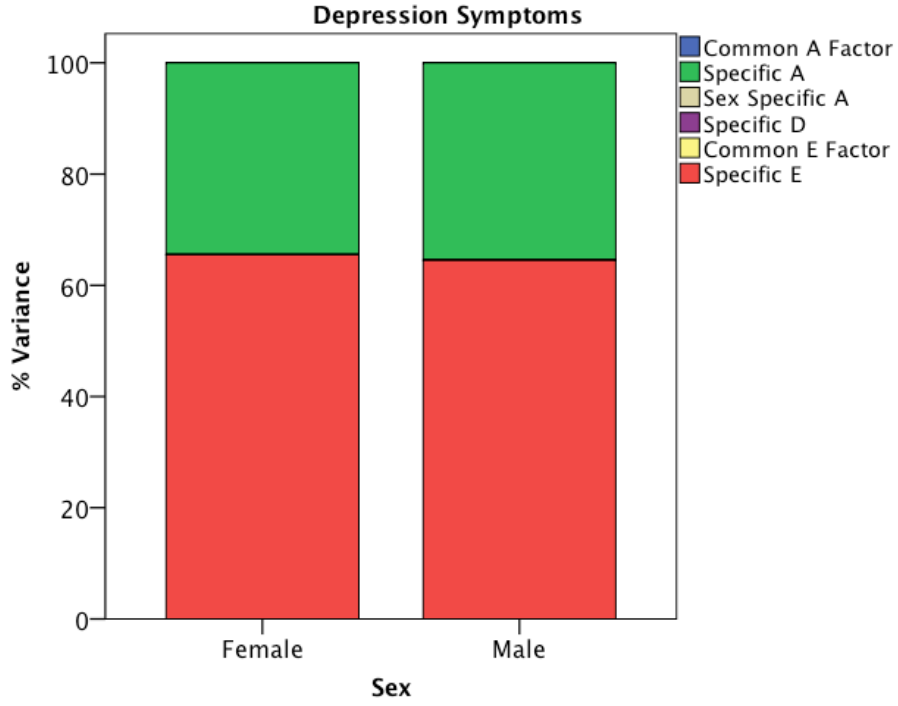


Figure 25: Proportion of variance in depression symptoms and impulsivity due to ADE components (Bivariate)



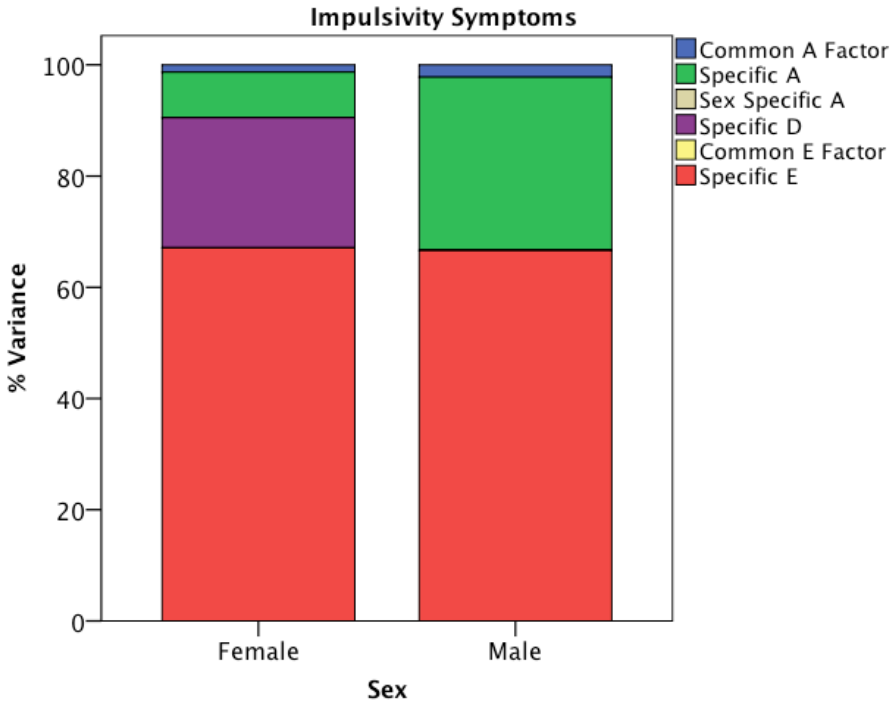


Table 42: Bivariate models of BMI, depression & impulsivity

able 1: Bivariate Sex Limitation BMI & Depression Symptoms (SCL-90)

								FEMALES:									MALES:											
lodel	**EP**	**-2LL**	**df**	**AIC**	**Diff LL**	**Difdf**	**p-value**	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**c11** **C11** [95%CI]	**c21** **C21** [95%CI]	**c22** **C22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**as11** **As11** [95%CI]	**as21** **As21** [95%CI]	**as22** **As22** [95%CI]	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**c11** **C11** [95%CI]	**c21** **C21** [95%CI]	**c22** **C22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**rA** - [95%CI]	**rC** - [95%CI]	**rE** - [95%CI]	**rgBMImf** - [95%CI]	**rgDEPmf** - [95%CI]
Full	25	68820.29	23827	21166.2	-	-	-	0.830 0.688 [0.394,0.912]	-0.030 -0.025 [-0.058,0.015]	0.569 0.325 [0.018,0.475]	0.096 0.009 [0,0.128]	-0.275 -0.026 [-0.102,0.068]	0.244 0.135 [0,0.281]	0.558 0.311 [0.292,0.332]	-0.004 -0.002 [-0.022,0.018]	0.965 0.931 [0.876,0.989]	0.583 0.339 [0.089,0.580]	0.126 0.073 [-0.001,0.156]	0.010 0.016 [0,0.339]	0.792 0.626 [0.525,0.671]	-0.032 -0.026 [-0.058,0.019]	0.618 0.383 [0.265,0.451]	0.014 0 [0,0.095]	0.000 0 [-0.035,0]	0.000 0 [0,0.092]	0.454 0.206 [0.187,0.227]	-0.003 -0.001 [-0.015,0.012]	0.816 0.666 [0.608,0.729]	-0.052 - [-0.117,0.042]	-0.748 - [-1,1]	-0.004 - [-0.041,0.033]	0.409 - [0.320,0.476]	0.488 - [0.125,0.500]
. Drop Sex specific effects	22	68829.4	23830	21169.4	9.12	3	0.03	0.977 0.955 [0.847,1.065]	-0.024 -0.023 [-0.06,0.012]	0.572 0.328 [0.157,0.482]	0.293 0.086 [0.001,0.187]	0.126 0.037 [-0.004,0.079]	0.359 0.145 [0.013,0.291]	0.562 0.316 [0.296,0.337]	0.002 0.001 [-0.018,0.021]	0.966 0.934 [0.881,0.977]	- - -	- - -	- - -	0.792 0.628 [0.585,0.673]	-0.026 -0.02 [-0.052,0.011]	0.618 0.383 [0.311,0.450]	0 0 [0,0.004]	0.002 0 [0,0.002]	0.005 0 [0,0.028]	0.457 0.209 [0.19,0.231]	0.002 0.001 [-0.004,0.014]	0.816 0.666 [0.608,0.730]	-0.042 - [-0.106,0.022]	0.331 - [-0.043,1]	0.003 - [-0.034,0.039]	0.5 - -	0.5 - -
drop as22 & as21	23	68825.08	23829	21167.08	4.79	2	0.09	0.833 0.694 [0.497,0.957]	-0.006 -0.005 [-0.031,0.043]	0.563 0.317 [0.145,0.526]	0 0 [0,0.053]	-0.392 0 [-0.019,0]	0.001 0.154 [0,0.3]	0.558 0.311 [0.292,0.331]	0.002 0.001 [-0.019,0.021]	0.967 0.936 [0.882,0.994]	0.585 0.342 [0.130,0.539]	- - -	- - -	0.79 0.624 [0.537,0.67]	-0.006 -0.005 [-0.033,0.033]	0.617 0.38 [0.253,0.45]	0.052 0.003 [0,0.084]	-0.032 -0.002 [-0.052,0]	0 0.001 [0,0.098]	0.454 0.206 [0.187,0.228]	0.002 0.001 [-0.013,0.014]	0.817 0.667 [0.609,0.734]	-0.01 - [-0.068,0.077]	-0.999 - [-1,1]	-0.002 - [-0.035,0.039]	0.409 - [0.347,0.479]	0.5 - -
1. AE	17	68830.05	23834	21162.05	9.77	7	0.2	0.833 0.694 [0.497,0.915]	-0.004 -0.003 [-0.032,0.025]	0.697 0.485 [0.428,0.543]	- - -	- - -	- - -	0.558 0.311 [0.292,0.331]	0.001 0.001 [-0.019 ,0.020]	0.96 0.921 [0.870,0.976]	0.585 0.342 [0.127,0.538]	- - -	- - -	0.792 0.627 [0.584,0.671]	-0.004 -0.003 [-0.027,0.021]	0.611 0.373 [0.307,0.440]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	0.001 0.0004 [-0.013,0.014]	0.822 0.676 [0.618,0.739]	-0.006 - [-0.055,0.043]	- - -	0.001 - [-0.035,0.037]	0.409 - [0.347,0.468]	0.5 - -
√. Drop e21 (rE)	15	68830.06	23835	21160.06	9.77	8	0.28	0.834 0.695 [0.498,0.915]	-0.004 -0.003 [-0.027,0.022]	0.697 0.485 [0.428,0.543]	- - -	- - -	- - -	0.558 0.311 [0.292,0.331]	- - -	0.96 0.921 [0.870,0.975]	0.585 0.342 [0.127,0.538]	- - -	- - -	0.792 0.627 [0.584,0.671]	-0.003 -0.003 [-0.023,0.018]	0.611 0.373 [0.307,0.440]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	- - -	0.822 0.676 [0.618,0.739]	-0.005 - [-0.048,0.037]	- - -	- - -	0.409 - [0.347,0.468]	0.5 - -
'. Drop a21 (rA)	13	68830.12	23836	21158.12	9.83	9	0.36	0.835 0.697 [0.501,0.917]	- - -	0.697 0.485 [0.428,0.543]	- - -	- - -	- - -	0.558 0.311 [0.292,0.331]	- - -	0.96 0.921 [0.870,0.975]	0.583 0.339 [0.126,0.534]	- - -	- - -	0.792 0.627 [0.584,0.671]	- - -	0.611 0.373 [0.306,0.440]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	- - -	0.822 0.676 [0.618,0.739]	- - -	- - -	- - -	0.41 - [0.349,0.469]	0.5 - -

le 2: Bivariate Sex Limitation BMI & Impulsivity (EPQ)

								FEMALES:									MALES:											
ɔdel	**EP**	**-2LL**	**df**	**AIC**	**Diff LL**	**Difdf**	**p-value**	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**d11** **D11** [95%CI]	**d21** **D21** [95%CI]	**d22** **D22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**as11** **As11** [95%CI]	**as21** **As21** [95%CI]	**as22** **As22** [95%CI]	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**d11** **D11** [95%CI]	**d21** **D21** [95%CI]	**d22** **D22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**rA** - [95%CI]	**rD** - [95%CI]	**rE** - [95%CI]	**rgBMImf** - [95%CI]	**rgIMPmf** - [95%CI]
ull	25	68401.95	23877	20647.95	-	-	-	0.87 0.757 [0.519,1.057]	-0.016 -0.014 [-0.090,0.043]	0.198 0.039 [0,0.190]	0.148 0.022 [0,0.187]	0.491 0.072 [0,0.213]	0.0002 0.241 [0.017,0.423]	0.557 0.311 [0.292,0.331]	0.034 0.019 [0,0.038]	0.907 0.824 [0.776,0.874]	0.507 0.257 [0,0.490]	-0.018 -0.009 [-0.151,0.115]	0.325 0.106 [0,0.321]	0.734 0.539 [0.366,0.653]	-0.043 -0.032 [-0.133,0.071]	0.551 0.305 [0.069,0.429]	0.296 0.088 [0,0.258]	0.282 0.084 [0,0.186]	0.0001 0.08 [0,0.325]	0.452 0.204 [0.185,0.226]	0.033 0.015 [0,0.050]	0.894 0.8 [0.734,0.873]	-0.079 - [-0.541,0.147]	0.999 - [-1,1]	0.037 - [0.001 ,0.074]	0.432 - [0.361,0.500]	0.259 - [0.000 ,0.500]
Drop Sex specific effects	22	68405.17	23880	20645.17	3.23	3	0.36	1.008 1.017 [0.908,1.075]	-0.021 -0.021 [-0.106,0.048]	0.246 0.061 [0,0.221]	0.142 0.02 [0,0.122]	0.501 0.071 [0,0.160]	0.275 0.327 [0.156,0.426]	0.558 0.311 [0.293,0.332]	0.034 0.019 [0,0.038]	0.906 0.822 [0.775,0.873]	- - -	- - -	- - -	0.672 0.451 [0.319,0.584]	-0.049 -0.033 [-0.122,0.057]	0.571 0.329 [0.133,0.426]	0.418 0.175 [0.051,0.306]	0.201 0.084 [0,0.175]	0.11 0.053 [0,0.256]	0.452 0.205 [0.186,0.226]	0.034 0.015 [0,0.050]	0.896 0.804 [0.737,0.877]	-0.085 - [-0.389 ,0.136]	0.877 - [-0.079,1]	0.038 - [0.001,0.075]	0.5 - [0.5 ,0.5]	0.497 - [0.436 ,0.500]
rop as22 & as21	23	68402.69	23879	20644.69	0.75	2	0.69	0.874 0.764 [0.533,0.764]	-0.014 -0.012 [-0.094,0.044]	0.223 0.05 [0.050,0.209]	0.108 0.012 [0,0.085]	0.582 0.063 [0,0.148]	0 0.338 [0.168,0.430]	0.557 0.311 [0.311,0.331]	0.034 0.019 [0.001,0.038]	0.906 0.821 [0.821,0.872]	0.511 0.261 [0.100,0.261]	- - -	- - -	0.739 0.546 [0.371,0.654]	-0.035 -0.026 [-0.126,0.071]	0.555 0.309 [0.081,0.431]	0.283 0.081 [0.080,0.254]	0.272 0.077 [0,0.178]	0 0.074 [0,0.074]	0.452 0.205 [0.204,0.205]	0.034 0.015 [0.001,0.015]	0.895 0.802 [0.802,0.874]	-0.063 - [-0.477,0.145]	0.999 - [-1,1]	0.038 - [0.001,0.075]	0.432 - [0.432,0.500]	0.498 - [0.398,0.500]
AE (drop d11,d21,d22)	17	68416.66	23884	20648.66	14.71	7	0.04	0.832 0.692 [0.499,0.908]	0.059 0.049 [0.023,0.076]	0.598 0.361 [0.311,0.411]	- - -	- - -	- - -	0.558 0.311 [0.292,0.331]	0.041 0.023 [0.004,0.041]	0.921 0.85 [0.803,0.900]	0.587 0.345 [0.134,0.536]	- - -	- - -	0.792 0.627 [0.584,0.671]	0.059 0.046 [0.022,0.072]	0.592 0.354 [0.281,0.354]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	0.04 0.018 [0.003,0.018]	0.911 0.832 [0.764,0.905]	0.098 - [0.046,0.152]	- - -	0.044 - [0.008,0.079]	0.408 - [0.348 ,0.466]	0.499 - [0.496,0.500]
.E/ADE AE/AE (keep df22)	18	68406.45	23883	20640.45	4.51	6	0.61	0.841 0.707 [0.513,0.925]	0.041 0.035 [0.014,0.060]	0.355 0.128 [0.036,0.266]	- - -	- - -	0.508 0.258 [0.111,0.367]	0.558 0.311 [0.293,0.332]	0.042 0.024 [0.006,0.042]	0.907 0.824 [0.777,0.874]	0.573 0.328 [0.117,0.521]	- - -	- - -	0.792 0.627 [0.585,0.672]	0.07 0.056 [0.026,0.087]	0.608 0.375 [0.302,0.448]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	0.042 0.019 [0.005,0.034]	0.899 0.811 [0.744,0.883]	0.115 - [0.053,0.178]	- - -	0.046 - [0.011,0.082]	0.413 - [0.353,0.471]	0.496 - [0.490,0.499]
Drop e21 (rE)	16	68413.11	23884	20645.11	11.17 6.66	7 1	0.13 0.01	0.846 0.716 [0.522,0.932]	0.059 0.05 [0.028,0.074]	0.38 0.148 [0.056,0.281]	- - -	- - -	0.488 0.238 [0.096,0.347]	0.558 0.311 [0.292,0.331]	- - -	0.908 0.824 [0.777,0.874]	0.567 0.321 [0.111,0.513]	- - -	- - -	0.792 0.628 [0.585,0.673]	0.093 0.074 [0.048,0.102]	0.603 0.373 [0.300,0.446]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	- - -	0.902 0.813 [0.746,0.886]	0.153 - [0.099,0.209]	- - -	0.415 - [0.356,0.473]	0.494 - [0.487,0.498]
Drop a21 (rA)	14	68444.25	23885	20674.25	42.3	8	< 0.0001	0.835 0.697 [0.501,0.917]	- - -	0.305 0.093 [0.011,0.237]	- - -	- - -	0.542 0.294 [0.141,0.399]	0.558 0.311 [0.292,0.331]	- - -	0.907 0.822 [0.776,0.873]	0.583 0.339 [0.126,0.534]	- - -	- - -	0.792 0.627 [0.584,0.671]	- - -	0.615 0.379 [0.306,0.451]	- - -	- - -	- - -	0.454 0.206 [0.187,0.227]	- - -	0.898 0.806 [0.740,0.879]	- - -	- - -	- - -	0.41 - [0.349,0.469]	0.5 - -

le 3: Bivariate Sex Limitation Depression Symptoms (SCL-90) & Impulsivity (EPQ)

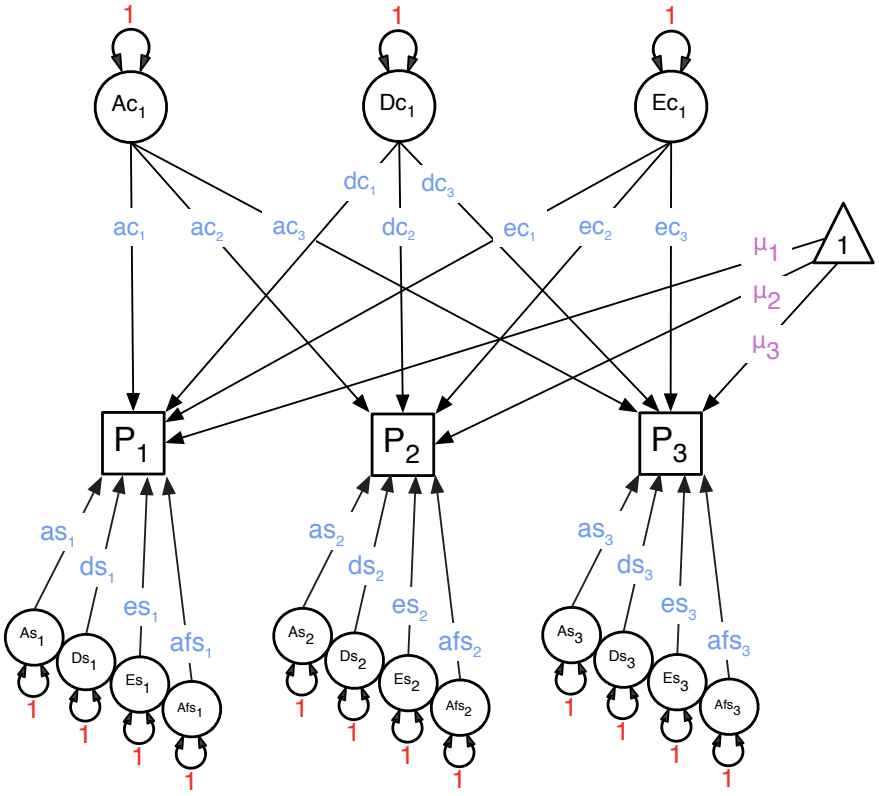
								FEMALES:									MALES:											
del	**EP**	**-2LL**	**df**	**AIC**	**Diff LL**	**Difdf**	**p-value**	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**d11** **D11** [95%CI]	**d21** **D21** [95%CI]	**d22** **D22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**as11** **As11** [95%CI]	**as21** **As21** [95%CI]	**as22** **As22** [95%CI]	**a11** **A11** [95%CI]	**a21** **A21** [95%CI]	**a22** **A22** [95%CI]	**d11** **D11** [95%CI]	**d21** **D21** [95%CI]	**d22** **D22** [95%CI]	**e11** **E11** [95%CI]	**e21** **E21** [95%CI]	**e22** **E22** [95%CI]	**rA** - [95%CI]	**rD** - [95%CI]	**rE** - [95%CI]	**rgDEPmf** - [95%CI]	**rgIMPmf** - [95%CI]
dl	25	67513.32	23906	19701.32	-	-	-	0.617 0.381 [0.127,0.428]	0.02 0.012 [-0.021,0.079]	0.247 0.062 [0,0.201]	0.034 0.001 [0,0.122]	-0.073 -0.002 [-0.119,0.074]	0.456 0.213 [0.36,0]	0.844 0.712 [0.673,0.754]	0.006 0.005 [-0.019,0.030]	0.847 0.718 [0.676,0.762]	-0.022 0.0005 [0,0.255]	-0.243 0.005 [-0.059,0.117]	-0.051 0.062 [0,0.300]	0.426 0.181 [0.064,0.342]	0.045 0.019 [-0.035,0.116]	0.571 0.329 [0.115,0.594]	0.357 0.127 [0,0.262]	-0.008 -0.003 [-0.105,0.051]	0.05 0.003 [0,0.225]	0.713 0.508 [0.462,0.560]	0.006 0.004 [-0.016,0.025]	0.857 0.701 [0.644,0.764]	0.078 - [-0.145,0.475]	-0.158 - [-1,1]	0.007 - [-0.026,0.042]	0.499 - [0.292,0.500]	0.352 - [0,0.500]
Drop Sex specific effects	22	67515.83	23909	19697.83	2.51	3	0.47	0.623 0.389 [0.219,0.404]	0.032 0.02 [-0.030,0.102]	0.267 0.072 [0,0.245]	0.034 0.001 [0,0.168]	-0.129 -0.004 [-0.095,0.051]	0.473 0.241 [0.082,0.378]	0.847 0.718 [0.662,0.766]	0.0004 0.0004 [-0.024,0.036]	0.86 0.74 [0.665,0.773]	- - -	- - -	- - -	0.424 0.18 [0.040,0.337]	0.067 0.029 [-0.043,0.119]	0.565 0.324 [0.061,0.410]	0.36 0.13 [0,0.292]	-0.014 -0.005 [-0.110,0.061]	0.051 0.003 [0,0.280]	0.713 0.509 [0.450,0.575]	0.0004 0.0003 [-0.020,0.030]	0.839 0.703 [0.630,0.783]	0.118 - [-0.195,0.595]	-0.264 - [-0.524,1]	0.0005 - [-0.034,0.050]	0.5 - [0.500,0.500]	0.495 - [0.347,0.500]
AE (drop d11,d21,d22)	16	67528.43	23914	19700.43 12.55	15.11 5	8 **0.03**	0.06	0.615 0.378 [0.334,0.423]	0.025 0.015 [-0.010,0.041]	0.559 0.313 [0.270,0.357]	- - -	- - -	- - -	0.847 0.717 [0.677,0.759]	0.007 0.006 [-0.018,0.030]	0.86 0.741 [0.700,0.784]	- - -	- - -	- - -	0.539 0.29 [0.239,0.343]	0.025 0.013 [-0.009,0.036]	0.554 0.308 [0.245,0.372]	- - -	- - -	- - -	0.725 0.526 [0.481,0.575]	0.006 0.005 [-0.016,0.025]	0.851 0.725 [0.666,0.789]	0.045 - [-0.029,0.119]	-	0.008 - [-0.025,0.040]	0.5 - [0.500,0.500]	0.5 - [0.497,0.500]
AE/ADE AE/AE (keep df22)	17	67516.58	23913	19690.58 II.	3.26 0.7	7 4	0.86 0.95	0.615 0.378 [0.334,0.423]	0.017 0.01 [-0.006,0.031]	0.29 0.084 [0.012,0.207]	- - -	- - -	0.503 0.253 [0.122,0.344]	0.847 0.717 [0.677,0.759]	0.008 0.006 [-0.017,0.029]	0.847 0.717 [0.676,0.760]	- - -	- - -	- - -	0.539 0.29 [0.239,0.343]	0.034 0.018 [-0.011,0.048]	0.573 0.33 [0.267,0.393]	- - -	- - -	- - -	0.726 0.526 [0.481,0.576]	0.008 0.006 [-0.014,0.025]	0.838 0.702 [0.645,0.765]	0.058 - [-0.036,0.154]	-	0.009 - [-0.023,0.041]	-	-
Drop e21 (rE)*	15	67516.89	23914	19688.89 II.	3.57 1.01	8 5	0.89 0.96	0.615 0.378 [0.334,0.423]	0.022 0.014 [0.001,0.030]	0.295 0.088 [0.015,0.209]	- - -	- - -	0.5 0.25 [0.120,0.284]	0.847 0.717 [0.677,0.759]	- - -	0.846 0.717 [0.676,0.760]	- - -	- - -	- - -	0.538 0.29 [0.238,0.342]	0.043 0.023 [0.001,0.047]	0.572 0.329 [0.266,0.393]	- - -	- - -	- - -	0.726 0.527 [0.482,0.576]	- - -	0.838 0.703 [0.646,0.766]	0.075 - [0.003,0.151]	-	-	-	-
rop e21 & drop a21 (rA)	13	67521.04	23915	19691.04 III.b IV.	7.72 4.46 4.15	9 2 1	0.56 0.11 **0.04**	0.615 0.378 [0.334,0.423]	- - -	0.285 0.081 [0.009,0.206]	- - -	- - -	0.506 0.256 [0.123,0.347]	0.847 0.717 [0.678,0.759]	- - -	0.846 0.716 [0.676,0.760]	- - -	- - -	- - -	0.539 0.29 [0.239,0.343]	- - -	0.574 0.33 [0.267,0.393]	- - -	- - -	- - -	0.725 0.526 [0.481,0.576]	- - -	0.838 0.702 [0.645,0.765]	- - -	- - -	- - -	- - -	- - -

Table 43: Trivariate independent pathway models of impulsivity, BMI & depression Sx

Table 44: Trivariate Cholesky parameterization of impulsivity, BMI & depression Sx

Figure 26: Trivariate independent pathway sex limitation model

a. Females



b. Males

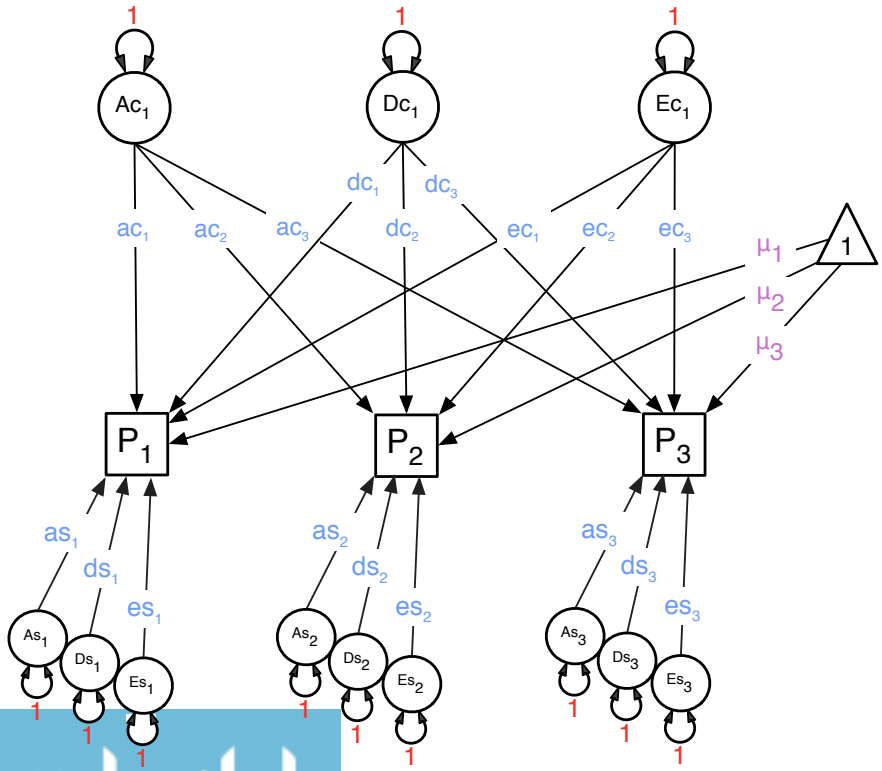
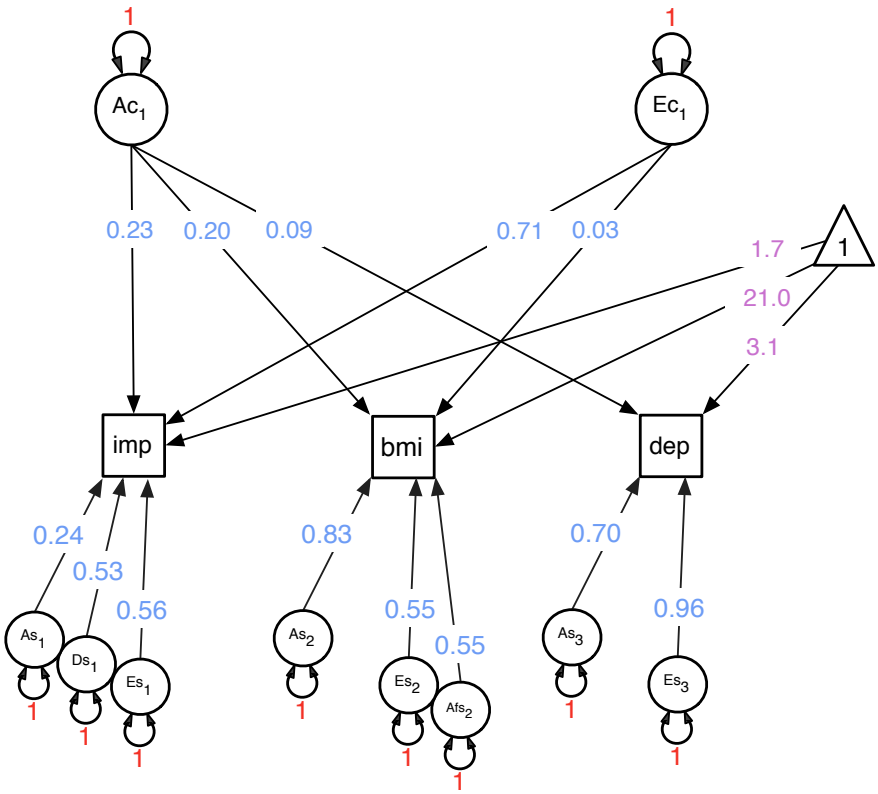


Figure 27: Best fitting model

a. Females



b. Males

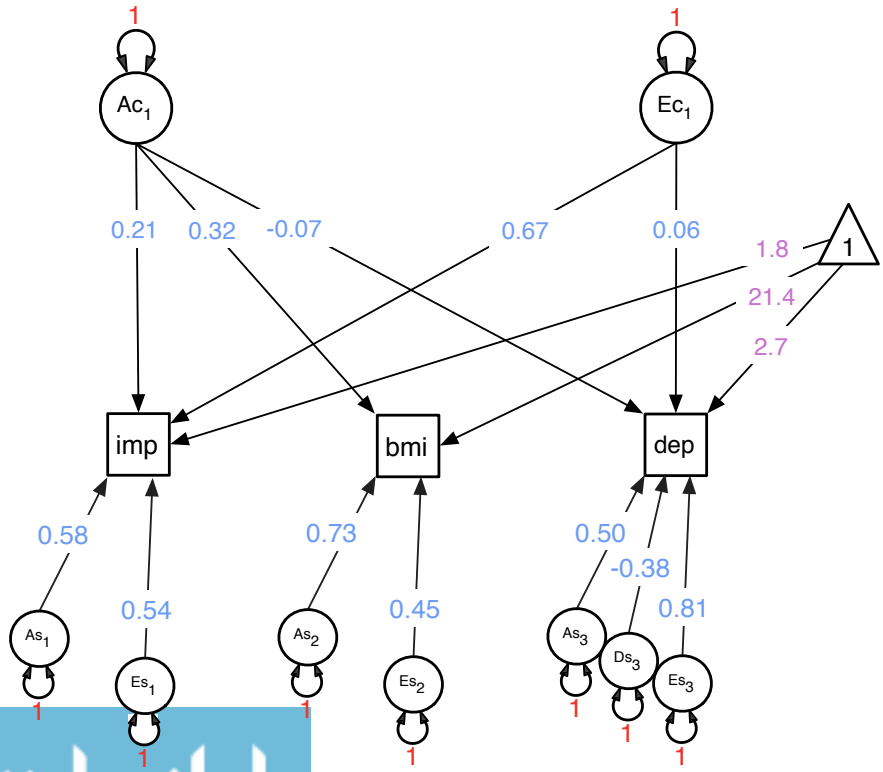


Figure 28: Proportion of variance in BMI accounted for by ADE components (Trivariate)

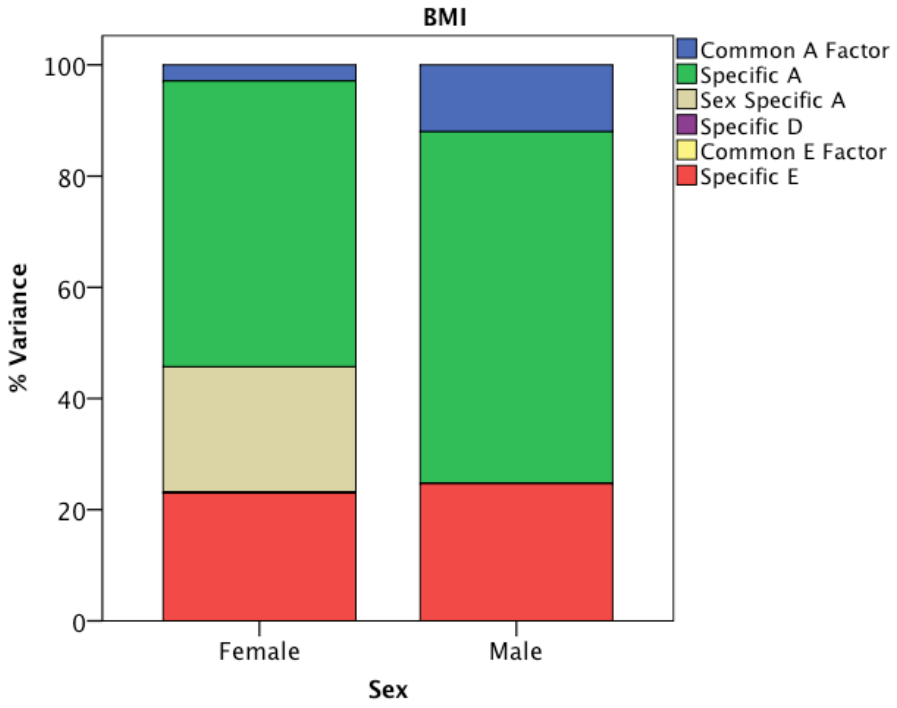


Figure 29: Proportion of variance in depression symptoms accounted for by ADE components (Trivariate)

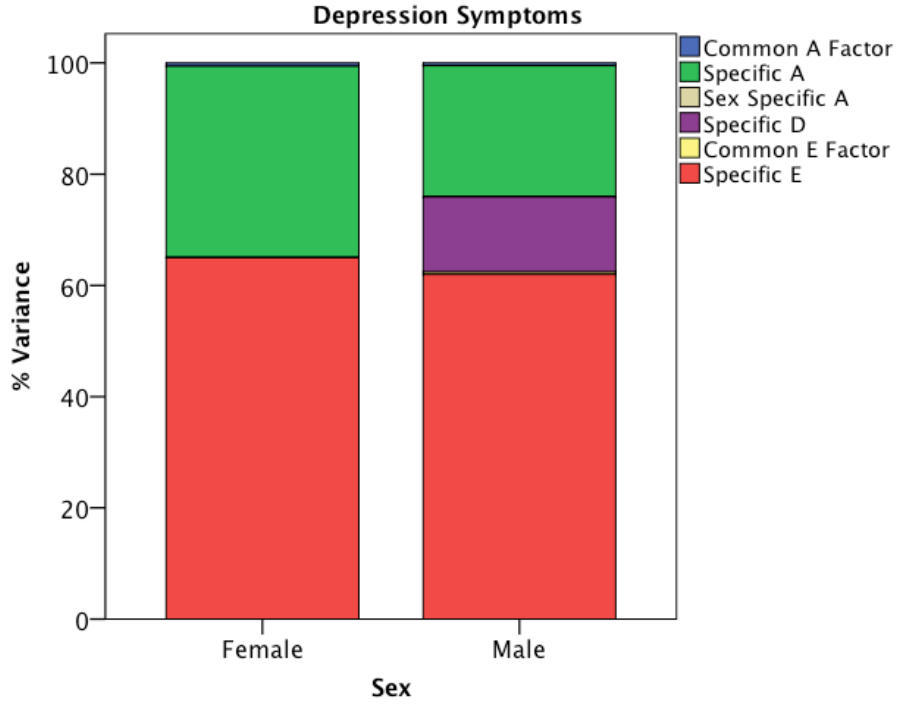
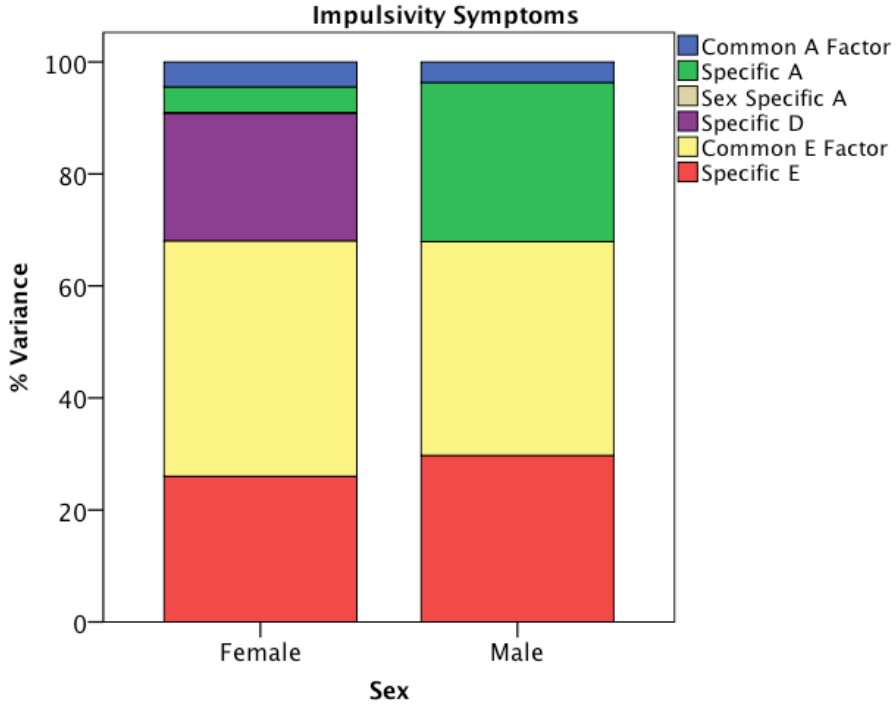


Figure 30: Proportion of variance in impulsivity symptoms accounted for by ADE components (Trivariate)



Chapter 8: Evidence of shared genetic risk between body composition and smoking behaviors

Adapted from:

1) On the genetic and environmental relationship of body mass index, smoking initiation and nicotine dependence in a population-based sample of twins. Roseann E. Peterson, Lindon J. Eaves, Hermine H. Maes. Presentation. XIX World Congress of Psychiatric Genetics, September 13th, 2011. Washington, D.C., USA.
2) Evidence of Shared Polygenic Risk Among Smoking Behaviors and Body Composition. Roseann E. Peterson, Xiangning (Sam) Chen, Jingchun Chen, Bradley T. Webb, Hermine H. Maes. Presentation. 4th International Conference on Quantitative Genetics, June 21, 2012. Edinburgh, Scotland, UK.

ABSTRACT

Obesity and nicotine dependence (ND) are complex, heterogeneous diseases, which pose a significant burden to public health, affecting 33 and 20 percent of Americans, respectively. Cross-sectional studies of ND are generally supportive of a negative relationship between smoking and body mass index (BMI), but a positive association is supported by the observation that within smoking cohorts, heavy smokers tend to be of increased body weight compared to light smokers. Genetic factors have consistently been demonstrated to influence individual differences in body mass index (BMI) and nicotine dependence (ND), with twin and family studies estimating heritabilities in the order of 0.70 and 0.60 respectively. A growing body of evidence demonstrates the utility of genome-wide association studies (GWAS) for identifying single nucleotide polymorphisms (SNP) that contribute to disease risk. The GWAS approach has been applied to BMI and smoking behaviors (SB) using sample sizes in the tens of thousands, yielding several putative risk variants of small effects on individual traits. However, most studies do not examine common versus specific genetic effects, despite many complex traits demonstrating comorbidity. Moreover, without consideration of genetically-correlated traits, the power of genome-wide studies of complex disease to detect etiologically relevant variation may be limited. Therefore, the purpose of this study was to investigate whether genetic variants affecting BMI or SB were common to multiple behaviors or were trait-specific. In total, 75 BMI and 54 SB associated SNPs were catalogued from large-scale GWAS meta-analyses and tested for association in n=2,802 (41% African-American) older adults (68-80 years old) from the Health Aging and Body Composition study (Heath ABC). Results indicated current smokers had significantly lower BMI and abdominal visceral fat than never or former smokers in both sexes. We observed three BMI-associated SNPs also nominally associated with smoking traits: rs1900273 in *STK33*, rs2145270 near *BMP2* and rs12127438 in the 1q42.2 locus. Additionally, three SB-associated SNPs were found to be nominally associated with body composition variables: rs11072774 in *CHRNB4*, rs2640732 in *SCARA3* and rs6945244 in *PDE1C*. These findings are suggestive of partially shared genetic risk between smoking and body composition. Future research should confirm these associations and address putative mechanisms underlying this overlapping genetic architecture.

INTRODUCTION

Obesity and nicotine dependence (ND) represent complex heterogeneous diseases affecting 33 and 20 percent of Americans, respectively (106, 127, 129). Both are associated with numerous medical conditions including cancer, cardiovascular disease and major depressive disorder (12, 127, 128, 366, 367). Phenotypic associations between smoking and body composition suggest a complex relationship and the causes of these associations remain incompletely understood. Cross-sectional studies of smoking behavior are typically supportive of a negative relationship between current smoking and body mass index (BMI) (109-111) which may be due in part to effects of nicotine on energy homeostasis including a reduction of energy intake (112-116) and enhanced capacity for energy expenditure (113, 368-370). Furthermore, the metabolic effects of nicotine might partially explain why smoking cessation is often followed by weight-gain (113, 117, 118). In contrast, however, a positive association is supported by the observation that within smoking cohorts, heavy smokers tend to be of increased body weight compared to light smokers (119-121). This may reflect a clustering of risky behaviors in addition to smoking- increased alcohol consumption, poor diet and reduced physical activity (371-374). Additionally, smoking has been associated with accumulation of visceral fat and increased waist circumference (122-124), which may be the result of nicotine's effects on sex hormones (375, 376) and cortisol levels (377, 378). For these reasons, the elucidation of the genetic and environmental mechanisms underlying these associations remains an important public health endeavor.

Genetic factors have consistently been demonstrated to influence individual differences in BMI and smoking behaviors (SB). Although an increase in energy intake coupled with reduced physical activity contributes to increases in adiposity, findings from twin and family studies have estimated large heritabilities on the order 0.70 for relative body weight (35, 36). Similarly, twin and family studies have estimated heritabilities in the order of 0.50-0.70 for smoking initiation and 0.60 for ND (379-383). To date, there have been no published multivariate twin and family studies on the genetic and environmental architecture of relative body weight and smoking behavior. However, our group has examined the possibility of shared genetic and environmental liability between BMI, smoking initiation and ND in a population-based sample of adult twins from the Virginia 30,000 study (n=14,177, 63.9% female). Preliminary results of fitting trivariate modified causal-contingent-common pathway models, which account for the contingency of ND on smoking initiation, found 1-5% of the variance in smoking initiation and nicotine dependence to be accounted for by genetic factors in common with BMI (Peterson *et al.*, in preparation). Preliminary results are presented in SUPPLEMENTAL MATERIAL section of this chapter.

Genome-wide association studies (GWAS) have successfully identified polymorphisms that contribute to disease risk for numerous complex traits and diseases (72). As applied to BMI and smoking behaviors (SB), GWAS have yielded several putative risk variants of small effects on individual traits using sample sizes in the tens of thousands. The first common single nucleotide polymorphisms (SNPs) associated with BMI and common obesity were in the *fat mass and obesity-associated* (*FTO*) gene and near *melanocortin 4 receptor* (*MC4R)* and have since been widely replicated (66, 130-

135). Additionally, two large-scale BMI meta-analyses by Thorleifsson *et al.* (2009) and Willer *et al*. (2009) yielded 13 genetic loci reaching genome-wide significance, including the previously implicated variants in and near *FTO* and *MC4R*. A recent mega-analysis, performed on 249,796 individuals from 46 studies has confirmed 32 BMI-associated SNPs. Although highly significant, the identified genetic variants had modest effects, corresponding to a 0.06-0.4 kg/m^2 change in BMI per allele, and modest odds ratios for obesity (BMI>30 kg/m^2) ranging between 1.03 and 1.3.

Similarly, large-scale GWAS for smoking traits have yielded putative risk variants of individually small effect. Three large meta-analyses of smoking initiation, consumption and cessation from Oxford-GlaxoSmithKline (Ox-GSK), the Tobacco and Genetics Consortium (TAG) and ENGAGE consortia were published as a series that included a combined analysis of over 140,000 individuals of European descent from 45 studies (384-386). Findings from these studies revealed one region associated with smoking initiation on 11p14.1, which includes the *brain-derived neurotrophic factor* (*BDNF)* (385). Additionally, the combined analysis of all three studies yielded five loci associated with smoking quantity, including the previously identified 15q25 locus which harbors three genes encoding neuronal nicotinic acetylcholine receptor subunits (NAChR), *CHRNA5*, *CHRNA3* and *CHRNA4* (384-386); a second locus encoding NAChRs on 8p11 in and near *CHRNB3* and *CHRNA6* (386); variants on 19q13 in and near *CYP2A6* and *CYP2B6* that code nicotine metabolizing enzymes (385, 386); SNPs on 7p14 in an intergenic region and variants on 10q25 in *LOC100188997*, a gene for non-coding RNA (385). A single variant near the *DBH* locus (9q34) was found to be associated with smoking cessation; this gene encodes dopamine β-hydroxylase, which catalyzes the conversion of dopamine to norepinephrine. As in the aforementioned studies of body composition, smoking variants were highly significant but had modest effect on behavior, with odds ratios ranging from 1.06 to 1.12 (384-386).

The causes of the observed associations between body composition and smoking behavior remain incompletely understood. It is possible that these traits share a common liability influenced by genetic and environmental factors. For example, genetic variants in *BDNF* have been associated with increased body mass and also with smoking initiation (63-65, 385). However, despite many complex traits demonstrating comorbidity, most studies do not examine common versus specific effects. Therefore, the purpose of this study was to investigate whether genetic variants affecting BMI or smoking behavior were common to multiple behaviors or were trait specific in n=2,802 (41% African-American) older community-dwelling adults (68-80 years old) from the Health Aging and Body Composition study (Health ABC). To the best of our knowledge, this is the first study to test BMI and SB variants in the same cohort across multiple traits.

METHODS

Participants

Participants were from the Health ABC study, a prospective community based sample of body composition changes over time in elderly American adults. Participants were recruited from 1997-1998 from Pittsburgh, PA, and Memphis, TN metropolitan area residents who were Medicare eligible and between the ages of 69 and 80 years old. Participants were excluded if they reported difficulty walking a quarter of a mile or climbing 10 stairs without resting. All participants gave written informed consent and both study sites approved the protocol. There were 1663 white and 1139 black participants included in the present study.

Phenotypes

BMI was calculated from laboratory measured height and weight during initial evaluation. To test various BMI thresholds, BMI was partitioned into clinical categories with BMI ranges of underweight <18, normal 18-25, overweight 25-30 and obese 30+ kg/m^2. Physical activity (PhyAct) was estimated from a structured interview of 27 questions and summarized as kcal/kg/week. Computerized tomography was used to determine abdominal visceral adiposity density (AbVFat). Smoking habits and race were self-reported via telephone interview. Smoking status (smoke) was defined as never, current or former smoker. Smoking status was further partitioned into ever smoker (EvSmo), former versus current smoker (cessation) and current smoker (CurSmo). Smoking duration was measured as pack years (PkYrs) and was calculated as the number of packs of cigarettes smoked per day multiplied by years as a smoker.

Genotyping

Genotyping in the Health ABC was performed by the Center for Inherited Disease Research using the Illumina Human 1M-Duo BeadChip system. Analysis was restricted to SNPs with minor allele frequency greater than or equal to 1%, call rate greater than or equal to 98% and Hardy-Weinberg Equilibrium p-value greater than 10^{-5}. There were 8 samples removed for genotypic sex mismatch.

Selection of SNPs

Preliminary SNP selection identified 78 variants meeting criteria for genome-wide or suggestive significance in either of two large meta-analyses of BMI; 43 from Thorleifsson *et al.* (2009) and 35 from Willer *et al.* (64, 65). Thorleifsson and colleagues report genome-wide significant ($p < 1.6^{-7}$) associations with 29 SNPs in 11 chromosomal regions, using a discovery sample of n=34,416 and replication sample of n=5,586. The Willer *et al.* meta-analysis detected 8 genome-wide significant ($p<5.0^{-8}$) SNPs in first- and second-stage samples of n=32,387 and n=54,316, respectively. The only variants found to be genome-wide significant in both meta-analyses were in and near *FTO* and *MC4R*. The remaining genetic loci were suggestive in the opposing meta-analyses

($p<0.05$), except rs7138803 on 12q13 ($p=0.14$). Significance level for one SNP, rs10938397 on 4p12, could not be compared between meta-analyses because there was no corresponding proxy SNP. Of the 78 BMI variants catalogued, 57 had matching SNPs on the Illumina Human 1M-Duo array. For the 20 SNPs not present, proxies (16 with $r^2>0.8$; 2 with $r^2>0.7$) were identified using SNP Annotation and Proxy Search (SNAP) V2.1 (147). Following removal of 2 variants from Willer *et al.* for which no proxies were available ($r^2>0.7$), a total of 75 SNPs remained.

SNP selection for smoking traits were catalogued from three large meta-analyses on smoking initiation, consumption and cessation from Ox-GSK, TAG and ENGAGE consortia which included a combined analysis of over 140,000 individuals from 45 studies (384-386). There were 510 SNPs reported in Ox-GSK associated with EvSmo, smoking quantity and cessation of which 157 remained significant ($p<10^{-5}$) in the combined sample with TAG and ENGAGE. The TAG consortium reported 5 SNPs associated with consumption, 8 with EvSmo and 1 with cessation ($p<10^{-8}$) in the combined sample. There were 921 SNPs reported by ENGAGE associated ($p<0.05$) with cigarette consumption and EvSmo of which 437 remained significant ($p<10^{-5}$) in the combined sample with TAG and Ox-GSK. There were a total of 595 SNPs catalogued from the three large meta-analyses that were significant at the $p<10^{-5}$ level in the combined analysis of which 179 appeared on the Health ABC Illumina Human 1M-Duo array. HapMap phase 2 (CEU, release 23, 90 individuals, 3.96 million SNPs) was used to determine independence of the 595 SNPs catalogued (70). SNP pruning at 0.7 level indicated 69 independent SNPs of which 54 appeared on the Health ABC Illumina Human 1M-Duo array. There were 15 SNPs catalogued from Ox-GSK from fine mapping of the 15q25 locus that did not have corresponding proxies on the Illumina array.

Haploview version 4.10 was used to determine phase and corresponding proxy alleles (148, 149). In order to avoid bias due to correlated effects, SNP pruning ($r^2>0.7$) was performed using PLINK v. 1.07p (150). In summary, there were 75 BMI and 54 SB SNPs used for association in this study. Although our SNP selection threshold was more liberal than the traditional genome-wide significance threshold, it was more conservative than other models of complex disease risk prediction (151, 152).

Association Analyses

Linear and logistic regression was used to incorporate effects of covariates on outcome variables of body composition and smoking traits. Given there are phenotypic and SNP allele frequency differences found in European and African ancestries, analyses were run separately for self-identified race, white and black. To reduce spurious associations due to population stratification, principal component (PC) scores reflecting ancestral population sub-structure of each subject were computed (192, 387, 388). Eigensoft (192, 193) was used to generate 10 PCs from 336,680 independent SNPs (linkage disequilibrium <0.5) in the European-American and 578,446 SNPs in the African-American sample. There were 12 participants removed from analyses due to outlying PC scores. PCs 1, 2 and 5 were associated with study variables and were therefore included as covariates in subsequent regression analyses along with gender and age. PLINK v. 1.07p was used for association and meta-analyses (150).

RESULTS

Phenotypic

Descriptive statistics for study variables are presented in Table 45. Analysis of variance indicated that males had significantly greater AbVFat ($F(1,2694)=88.88$, $p=8.71\times10^{-21}$) and longer duration of tobacco exposure as assessed by PkYrs ($F(1,2758)=185.22$, $p=7.11\times10^{-41}$) than females. There were no statistical differences on PhyAct by gender. Pearson's Chi-Square and subsequent *post hoc* analyses indicated that males were more likely to be former smokers and females more likely to have never smoked (chi-squ=236.5, $p=4.0\times10^{-52}$). Additionally, females were more likely to be obese (chi-squ=24.2, $p=8.2\times10^{-7}$). As depicted in Figure 31, current smokers had significantly lower BMI than never or former smokers in males ($F(2,1362)=18.9$, $p=7.83\times10^{-9}$) and in females ($F(2,1430)=13.15$, $p= 2.17\times10^{-6}$). Similarly, current smokers had significantly lower AbVFat than never or former smokers in males ($F(2,1309)=20.60$, $p=1.54\times10^{-9}$) and females ($F(2,1377)=10.01$, $p=4.8\times10^{-5}$)(Figure 32). As shown in Figure 33, there were no significant differences in PkYrs across BMI categories in males ($F(3,1342)=1.45$, $p=0.330$). However, in females the underweight group had significantly greater PkYrs than the normal, overweight and obese groups ($F(3,1410)=5.75$, $p=0.001$).

BMI SNPs

Among genetic variants previously implicated in BMI, 23 were associated ($p<0.05$) with either BMI, AbVFat, BMICat or obesity (Table 48). Twelve of which were in the same direction for both racial groups. Table 46 lists association results suggestive for multiple traits. There were three BMI SNPs nominally associated with both body composition variables and smoking traits. The top associated SNP was rs1900273 on chromosome 11 in STK33 for association with BMI ($p=0.001$). This SNP was also associated with AbVFat and PkYrs in both samples ($p<0.023$). SNP rs2145270 near BMP2 on chromosome 20 was associated with obesity ($p=0.014$), smoking status ($p=0.007$) and EvSmo ($p=0.016$). Finally, rs12127438 in the 1q42.2 locus was associated with BMI and BMICat in both ethnicity-based cohorts ($p<0.033$) and PkYrs in the European-American group ($p=0.036$).

Smoking SNPs

Among genetic variants previously implicated in SB, 13 were associated ($p<0.05$) with either Smoke, PkYrs, EvSmo, cessation or CurSmo (Table 49). Seven of which were in the same direction for both racial groups. The top associated SB SNP was rs9633423 on chromosome 1 with smoking cessation ($p=0.008$). Table 47 lists association results suggestive in multiple traits for previously implicated SB SNPs. There were three SB SNPs associated with both body composition variables and smoking traits. SNP rs11072774 on chromosome 15 in CHRNB4 was associated with BMI ($p=0.037$), obesity ($p=0.003$) and CurSmo($p=0.020$) in both racial groups. In the white group, rs2640732 on chromosome 8 in SCARA3 was significantly associated with cessation, CurSmo and obesity ($p<0.033$). SNP rs6945244 on chromosome 7 in PDE1C was associated with

Smoke, EvSmo and PhyAct in both groups and additionally with obesity in the white group ($p<0.037$).

DISCUSSION

The purpose of this work was to examine phenotypic associations between body composition and smoking behavior in an elderly cohort and to test if genetic variants shown previously to be associated with either body composition or smoking behavior were associated with multiple traits in the Health ABC study. Since studies report significant phenotypic associations between body composition and smoking behavior, the work presented here investigated whether the association of these traits was due in part to shared genetic liability. Phenotypic results from the Health ABC study indicated there were body composition differences between smoking status groups. Specifically, current smokers tended to have lower BMI and abdominal visceral fat than former or non-smokers. These results are in agreement with other cross-sectional studies of smoking behavior, which are supportive of lower body weight in smokers (109-111).

To examine shared genetic liability, 75 BMI and 54 SB variants catalogued from large-scale GWAS meta-analyses were tested for association with body composition and smoking behavior variables. Among genetic variants previously implicated in BMI, 23 were nominally associated ($p<0.05$) in this sample with BMI, abdominal visceral fat or obesity in the expected direction, which included SNPs in or near *FTO* and *MC4R*. Among these, there were three variants that were also nominally associated with smoking traits in the Health ABC study. The first SNP was in the 1q42.2 locus between the *TSNAX* and *DISC1* genes and was negatively associated with BMI and pack years. This suggests that a BMI-decreasing allele is also associated with decreased smoking duration. However, a SNP in the STK33 gene (11p15.4), was found to be negatively associated with abdominal visceral fat and BMI but positively associated with pack years, suggesting that this allele is associated with lower body weight but with increased smoking duration. A third variant, on chromosome 20p12.3, was associated with a decrease in obesity but an increase in ever smoking with the closest gene *bone morphogenetic protein 2* (*BMP2*).

Among genetic variants previously implicated in smoking behavior, 13 were nominally associated ($p<0.05$) with smoking variables in the Health ABC study. Of these, three were also nominally associated with body composition variables. The first variant, on chromosome 7 located within the gene *PDE1C*, was associated with never smoking, decreased obesity and increase in physical activity. A second SNP, on chromosome 8 in the gene *SCARA3*, was associated with cessation, non-smoking and an increase in obesity. The third variant, located at the 15q25 locus, resides near a cluster of genes encoding nicotinic acetylcholine receptors and found to be associated with increased BMI and obesity, as well as with former and non-smoking. The genetic associations found in the Health ABC study reflect the complex phenotypic associations found between these traits. Although SNPs in and near *BDNF* were previously associated with body composition and smoking behaviors (64, 65, 385), there was no evidence of association in the Health ABC sample. However, despite their preliminary nature, these results merit future research and, in particular, follow-up in additional replication cohorts.

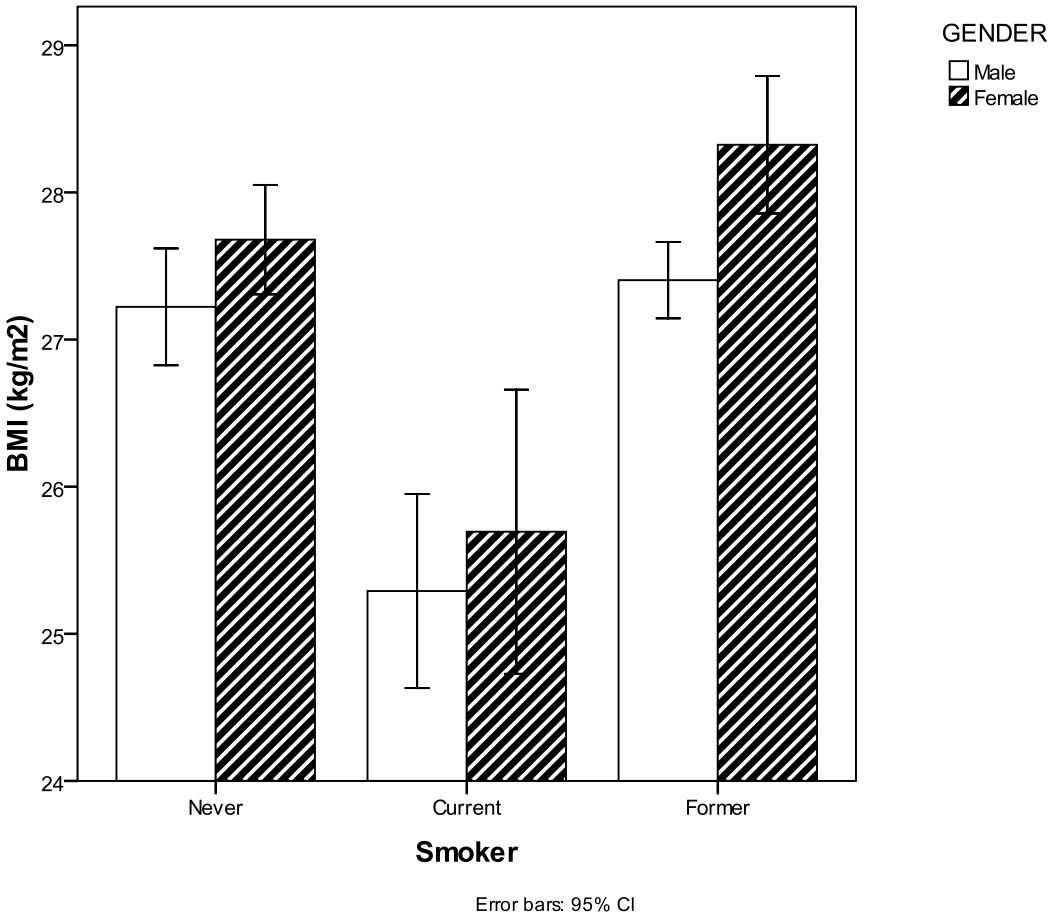
It is important to consider that genetic variants selected for association in this study, although significant in the meta-analyses from which they were catalogued, demonstrated relatively modest effects on their respective traits. As a result, replication attempts will have limited power to achieve genome-wide significance (125). Issues related to multiple testing further complicate this. The likelihood of observing a false positive finding increases with the number of tests performed and significance values reported here were not corrected for multiple testing. However, the results from Health ABC are preliminary, and several additional studies will be incorporated into the final analyses, with significance evaluated by appropriate measures including Bonferroni correction and empirical significance derived by permutation procedures.

Interpretation of these results should consider several limitations. First, this study was conducted using a selected sample. That is, participants were eligible if they were both elderly and in relatively good physical health. It is possible that this ascertainment strategy influenced the results and limits the ability to generalize the findings across the lifespan. Additionally, sex differences on the genetic analyses of these traits were not assessed. Further studies are warranted to determine effects of age and gender on the genetics of body composition, smoking behaviors and the causes of correlation between these traits.

Preliminary results were suggestive of partially shared genetic risk between smoking and body composition. Without consideration of genetically-correlated traits, genome-wide studies of complex disease may be limited in their power to detect etiologically relevant variation. Future research needs to address mechanisms underlying the associations between these traits and moderating effects of the environment to aid both obesity and nicotine dependence prevention and treatment efforts.

TABLES AND FIGURES

Figure 31: BMI by smoking status in males and females from the HABC study



Note: BMI = body mass index, kg = kilograms, m = meter.

Figure 32: Mean abdominal visceral fat density by smoking status in males and females

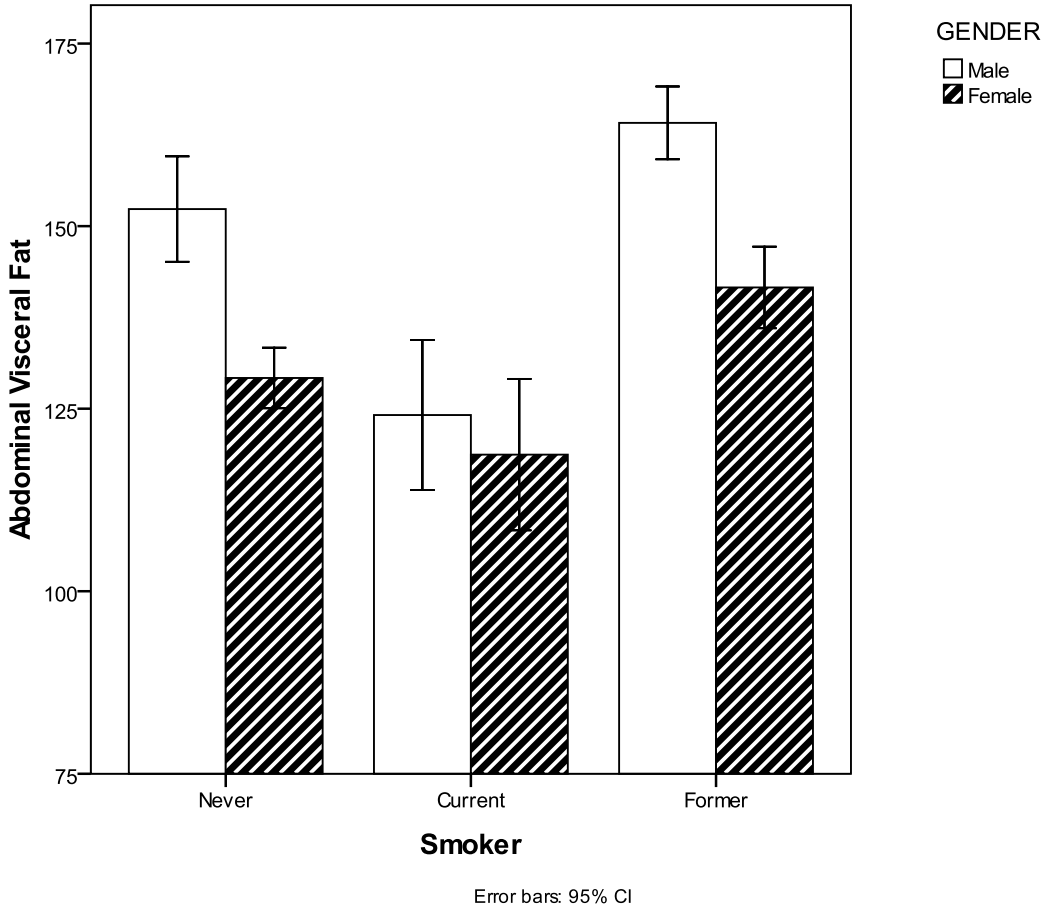
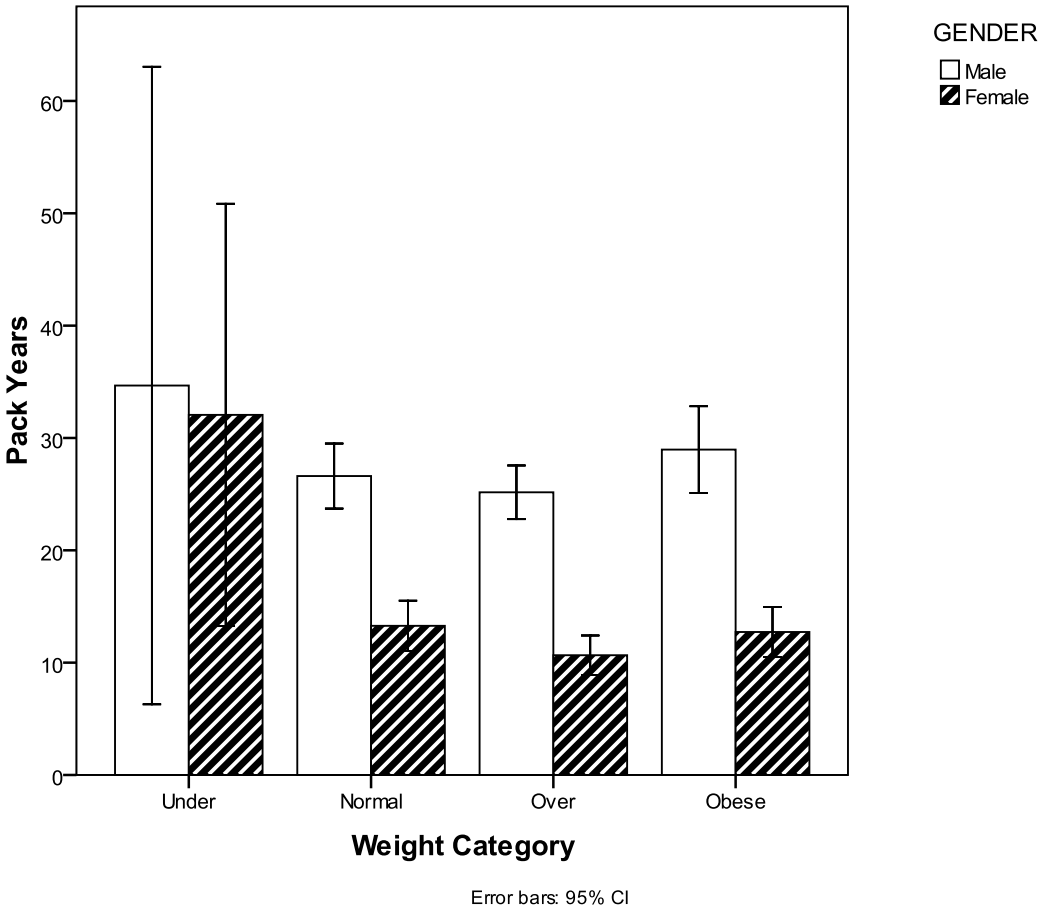


Figure 33: Mean pack years by BMI category in males and females



Note: BMI = body mass index, kg = kilograms, m = meter.

Table 45: Descriptive statistics for HABC study variables by gender

	Overall	Males	Females
N	2802	1367	1435
(%)		(48.8%)	(51.2%)
Race			
White	1663	879	784
Black	1139	488	651
Age (yrs) mean	73.6	73.8	73.5
BMI (kg/m2) mean	27.4	27.1	27.7
AbVFat mean	144	156.3	132.3
PhyAct (kcal/kg/wk) mean	82.8	81.7	83.9
Obese N	715	292	423
(%)	(25.5%)		
Smoke N			
Never	1206	393	813
(%)	(43%)		
Current	293	150	143
(%)	(10.5%)		
Former	1299	822	477
(%)	(46.4%)		
Pack Years mean	19.2	26.5	12.4
(median)	4	17	0

Note: BMI = body mass index, kg = kilograms, m = meter, kcal = kilocalories, wk = week, AbVFat = abdominal visceral fat, PhyAct = physical activity.

Table 46: Association results for SNPs previously implicated in BMI suggestive for multiple traits

				White			Black			Meta-Analysis	
Chr	SNP	A1	Trait	β/OR	SE	P	β/OR	SE	P	β/OR	P
1	rs12127438	G	BMI	-0.034	0.02	0.160	-0.048	0.03	0.101	-0.040	**0.033**
			BMICat	-0.056	0.03	**0.033**	-0.030	0.04	0.460	-0.048	**0.029**
			PkYrs	-0.051	0.02	**0.036**	0.046	0.03	0.112	-0.004	0.942
11	rs1900273	C	AbVFat	-0.041	0.02	0.091	-0.052	0.03	0.088	-0.045	**0.017**
			BMI	-0.057	0.02	**0.020**	-0.062	0.03	**0.031**	-0.059	**0.001**
			BMICat	-0.038	0.02	0.120	-0.066	0.03	0.056	-0.048	**0.017**
			PkYrs	0.056	0.02	**0.021**	0.023	0.03	0.425	0.043	**0.023**
20	rs2145270	C	EvSmo	1.171	0.08	**0.038**	1.121	0.09	0.201	1.150	**0.016**
			Obesity	0.851	0.09	0.085	0.852	0.09	0.080	0.851	**0.014**

Table 47: Association results for SNPs previously implicated in smoking behaviors suggestive for multiple traits

				White			Black			Meta-Analysis	
Chr	SNP	A1	Trait	β/OR	SE	P	β/OR	SE	P	β/OR	P
7	rs6945244	T	EvSmo	0.882	0.08	0.094	0.883	0.10	0.191	0.882	**0.034**
			Obesity	0.821	0.09	**0.033**	0.986	0.10	0.885	0.897	0.238
			PhyAct	0.024	0.02	0.323	0.062	0.03	**0.037**	0.040	**0.037**
8	rs2640732	G	Cessation	0.714	0.16	**0.033**	1.306	0.15	0.080	0.967	0.911
			CurSmo	0.716	0.15	**0.028**	1.216	0.14	0.155	0.937	0.805
			Obesity	1.228	0.09	**0.025**	0.930	0.11	0.516	1.077	0.595
15	rs11072774	T	BMI	0.024	0.02	0.315	0.059	0.03	**0.041**	0.039	**0.037**
			Cessation	0.589	0.22	**0.018**	0.870	0.18	0.426	0.733	0.109
			CurSmo	0.641	0.22	**0.041**	0.799	0.16	0.168	0.739	**0.020**
			Obesity	1.207	0.12	0.102	1.353	0.12	**0.011**	1.276	**0.003**

Note: Chr = chromosome, A1 = allele tested, β = beta estimate for linear regression, OR = odds ratio for logistic regression, SE = standard error of the estimate, SNP = single nucleotide polymorphism, BMI = body mass index, AbVFat = abdominal visceral fat, PhyAct = physical activity, BMICat = clinical BMI category (under, normal, overweight, obese), Smoke = smoking status (never, current, former), PkYrs = pack years, EvSmo = ever vs never smoked, Cessation = former vs current smoker, CurSmo = current smoker.

Table 48: Association results for SNPs previously implicated in BMI

				White			Black			Meta-Analysis	
Chr	SNP	A1	Trait	β/OR	SE	P	β/OR	SE	P	β/OR	P
1	rs3766431	T	AbVFat	0.017	0.02	0.477	-0.062	0.03	**0.040**	-0.020	0.608
1	rs9424977	C	PhyAct	0.018	0.02	0.458	0.077	0.03	**0.010**	-0.028	0.559
1	rs3101336	A	Cessation	0.718	0.17	**0.045**	0.971	0.12	0.811	0.851	0.281
			CurSmo	0.713	0.16	**0.033**	0.969	0.11	0.783	0.848	0.277
			PhyAct	0.036	0.03	0.153	-0.062	0.03	**0.037**	-0.012	0.810
1	rs2568958	G	Cessation	0.715	0.17	**0.043**	0.970	0.12	0.808	0.849	0.281
			CurSmo	0.711	0.16	**0.032**	0.969	0.11	0.786	0.847	0.279
			PhyAct	0.035	0.02	0.159	-0.062	0.03	**0.037**	-0.012	0.804
1	rs2815752	C	Cessation	0.716	0.17	**0.044**	0.970	0.12	0.808	0.850	0.281
			CurSmo	0.711	0.16	**0.032**	0.969	0.11	0.786	0.847	0.279
			PhyAct	0.036	0.02	0.153	-0.062	0.03	**0.037**	-0.012	0.810
1	rs1973993	T	AbVFat	-0.047	0.02	0.053	0.063	0.03	**0.041**	-0.053	**0.005**
1	rs10783050	C	AbVFat	0.071	0.02	**0.003**	0.015	0.03	0.635	0.047	0.097
			BMI	0.055	0.02	**0.024**	0.014	0.03	0.642	0.038	0.060
1	rs10913469	C	BMI	0.049	0.02	**0.045**	-0.013	0.03	0.648	0.020	0.523
1	rs12127438	G	BMI	-0.034	0.02	0.160	-0.048	0.03	0.101	-0.040	**0.033**
			BMICat	-0.056	0.03	**0.033**	-0.030	0.04	0.460	-0.048	**0.029**
			PkYrs	-0.051	0.02	**0.036**	0.046	0.03	0.112	-0.004	0.942
2	rs2867125	A	AbVFat	0.000	0.02	0.993	0.066	0.03	**0.029**	0.031	0.349
			Obesity	0.716	0.13	**0.009**	1.093	0.14	0.520	0.882	0.552
			PhyAct	-0.043	0.02	0.082	0.079	0.03	**0.008**	0.017	0.785
2	rs4854344	G	Obesity	0.716	0.13	**0.010**	0.970	0.11	0.774	0.840	0.249
			PhyAct	-0.042	0.02	0.089	0.087	0.03	**0.004**	0.021	0.743
2	rs7561317	A	Obesity	0.715	0.13	**0.009**	1.004	0.10	0.973	0.854	0.353
			PhyAct	-0.044	0.02	0.080	0.091	0.03	**0.002**	0.023	0.738
2	rs10206343	C	BMI	-0.056	0.02	**0.020**	-0.009	0.03	0.763	-0.035	0.139
3	rs7647305	T	BMI	-0.034	0.02	0.159	-0.053	0.03	0.066	-0.042	**0.024**
			BMICat	-0.024	0.03	0.426	-0.078	0.03	**0.024**	-0.049	0.065
5	rs467650	C	Obesity	0.971	0.09	0.755	0.807	0.09	**0.020**	0.884	0.182
6	rs1524097	C	EvSmo	0.708	0.11	**0.002**	0.894	0.12	0.366	0.790	**0.042**
			PkYrs	-0.051	0.02	**0.035**	0.016	0.03	0.586	-0.020	0.559
			Smoke	-0.154	0.05	**0.001**	-0.057	0.05	0.283	-0.109	**0.026**
7	rs7810507	A	AbVFat	0.044	0.02	0.068	0.055	0.03	0.071	0.048	**0.011**
			PkYrs	0.003	0.02	0.896	0.092	0.03	**0.002**	0.046	0.298
8	rs17069257	C	BMI	0.053	0.02	**0.029**	-0.018	0.03	0.542	0.019	0.582
			BMICat	0.069	0.03	**0.039**	-0.043	0.04	0.289	0.015	0.785
			Cessation	0.718	0.22	0.133	1.440	0.15	**0.015**	1.036	0.918
			Smoke	-0.014	0.04	0.738	-0.095	0.05	**0.040**	-0.053	0.189
9	rs4742700	A	EvSmo	1.234	0.08	**0.008**	1.040	0.12	0.743	1.161	0.069
			Smoke	0.082	0.03	**0.015**	0.020	0.05	0.696	0.063	**0.025**
9	rs867559	G	Cessation	0.762	0.22	0.211	1.378	0.13	**0.016**	1.051	0.868
			Smoke	-0.047	0.04	0.268	-0.078	0.04	0.056	-0.063	**0.032**
10	rs11255232	G	BMICat	0.079	0.03	**0.023**	-0.006	0.08	0.939	0.066	**0.038**
			Obesity	1.405	0.12	**0.004**	0.985	0.23	0.946	1.239	0.207
11	rs1900273	C	AbVFat	-0.041	0.02	0.091	-0.052	0.03	0.088	-0.045	**0.017**
			BMI	-0.057	0.02	**0.020**	-0.062	0.03	**0.031**	-0.059	**0.001**
			BMICat	-0.038	0.02	0.120	-0.066	0.03	0.056	-0.048	**0.017**

			PkYrs	0.056	0.02	**0.021**	0.023	0.03	0.425	0.043	**0.023**
11	rs7481311	T	AbVFat	-0.020	0.02	0.424	-0.086	0.03	**0.005**	-0.051	0.126
11	rs10835211	A	PkYrs	-0.007	0.02	0.793	-0.059	0.03	**0.043**	-0.030	0.243
11	rs4752856	A	Cessation	1.409	0.15	**0.027**	0.711	0.22	0.126	1.020	0.955
			Obesity	1.173	0.09	0.089	0.731	0.16	**0.049**	0.942	0.800
12	rs7138803	A	BMI	-0.027	0.02	0.264	0.068	0.03	**0.018**	0.019	0.685
13	rs7336332	G	Cessation	0.893	0.23	0.630	1.377	0.14	**0.018**	1.157	0.493
			CurSmo	0.827	0.23	0.414	1.281	0.12	**0.046**	1.076	0.734
15	rs12324805	C	AbVFat	-0.057	0.02	**0.018**	0.032	0.03	0.300	-0.015	0.739
			PhyAct	0.067	0.02	**0.007**	0.004	0.03	0.903	0.038	0.234
15	rs8024593	G	PhyAct	-0.051	0.02	**0.040**	0.022	0.03	0.477	-0.039	**0.040**
16	rs6499640	G	BMICat	-0.010	0.02	0.699	-0.069	0.03	**0.046**	-0.034	0.239
16	rs8050136	A	BMI	0.058	0.02	**0.017**	-0.030	0.03	0.291	0.015	0.730
			BMICat	0.065	0.03	**0.010**	-0.016	0.03	0.642	0.028	0.490
			Obesity	1.232	0.09	**0.023**	0.910	0.09	0.303	1.059	0.707
16	rs3751812	T	BMI	0.055	0.02	**0.025**	-0.036	0.03	0.219	0.011	0.813
			BMICat	0.062	0.03	**0.016**	-0.041	0.06	0.476	0.023	0.640
			Obesity	1.229	0.09	**0.025**	0.744	0.16	0.069	0.974	0.915
16	rs11075989	T	BMI	0.057	0.02	**0.020**	-0.012	0.03	0.682	0.024	0.480
			BMICat	0.063	0.03	**0.014**	0.003	0.03	0.934	0.037	0.217
			Obesity	1.229	0.09	**0.025**	0.949	0.09	0.566	1.080	0.553
16	rs7190492	A	BMI	-0.054	0.02	**0.029**	0.035	0.03	0.225	-0.011	0.812
16	rs8044769	T	BMI	-0.048	0.02	**0.048**	0.034	0.03	0.244	-0.009	0.834
			BMICat	-0.052	0.03	**0.040**	0.050	0.04	0.189	-0.005	0.927
18	rs10871777	G	BMI	0.041	0.02	0.092	0.045	0.03	0.118	0.043	**0.022**
			PhyAct	0.056	0.02	**0.024**	-0.021	0.03	0.477	0.019	0.619
18	rs12970134	A	BMI	0.049	0.02	**0.044**	0.052	0.03	0.071	0.050	**0.007**
20	rs2145270	C	EvSmo	1.171	0.08	**0.038**	1.121	0.09	0.201	1.150	**0.016**
			Obesity	0.851	0.09	0.085	0.852	0.09	0.080	0.851	**0.014**
			Smoke	0.071	0.03	**0.030**	0.061	0.04	0.108	0.067	**0.007**
22	rs4823535	A	PhyAct	-0.019	0.02	0.439	0.085	0.03	**0.004**	0.032	0.544

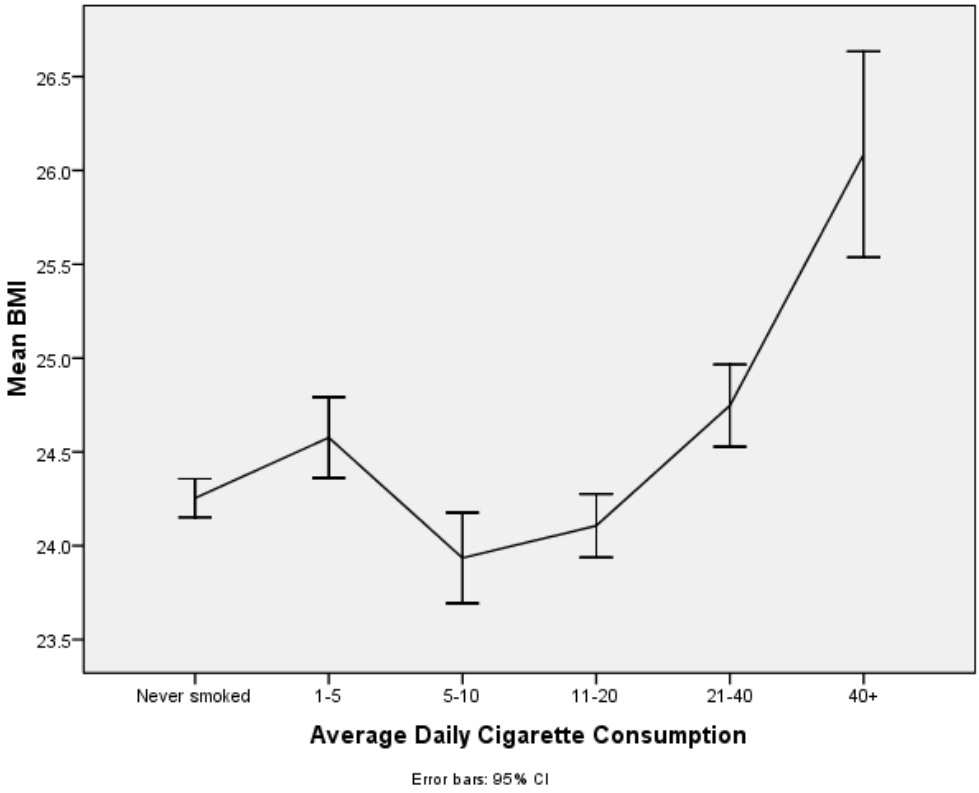
Note: Chr = chromosome, A1 = allele tested, β = beta estimate for liner regression, OR = odds ratio for logistic regression, SE = standard error of the estimate, SNP = single nucleotide polymorphism, BMI = body mass index, AbVFat = abdominal visceral fat, PhyAct = physical activity, BMICat = clinical BMI category (under, normal, overweight, obese), Smoke = smoking status (never, current, former), EvSmo = ever smoked, Cessation = former vs current smoker, CurSmo = current smoker.

Table 49: Association results for SNPs previously implicated in smoking behaviors

				White			Black			Meta-Analysis	
Chr	SNP	A1	Trait	β/OR	SE	P	β/OR	SE	P	β/OR	P
1	rs839758	G	BMI	-0.011	0.02	0.659	0.061	0.03	**0.035**	0.023	0.514
			BMICat	0.005	0.02	0.839	0.071	0.03	**0.029**	0.035	0.289
			Obesity	1.002	0.09	0.980	1.268	0.09	**0.008**	1.128	0.307
1	rs2782641	G	Cessation	1.078	0.15	0.626	0.748	0.13	**0.031**	0.891	0.529
1	rs10888740	A	PkYrs	-0.012	0.02	0.617	-0.059	0.03	**0.044**	-0.033	0.159
1	rs9633423	T	Cessation	1.299	0.16	0.092	1.340	0.14	**0.043**	1.321	**0.009**
			CurSmo	1.283	0.15	0.096	1.194	0.13	0.161	1.230	**0.032**
			PhyAct	-0.018	0.02	0.471	0.067	0.03	**0.025**	0.023	0.589
1	rs6683734	A	AbVFat	-0.015	0.02	0.536	0.074	0.03	**0.016**	-0.042	0.156
			PhyAct	-0.049	0.02	**0.047**	0.033	0.03	0.276	-0.010	0.807
2	rs16824949	G	PkYrs	0.036	0.02	0.143	-0.067	0.03	**0.025**	-0.014	0.782
7	rs6945244	T	EvSmo	0.882	0.08	0.094	0.883	0.10	0.191	0.882	**0.034**
			Obesity	0.821	0.09	**0.033**	0.986	0.10	0.885	0.897	0.238
			PhyAct	0.024	0.02	0.323	0.062	0.03	**0.037**	0.040	**0.037**
			Smoke	-0.057	0.03	0.079	-0.075	0.04	0.065	-0.064	**0.011**
7	rs6948856	A	Obesity	0.761	0.11	**0.012**	1.086	0.09	0.384	0.913	0.608
8	rs2640732	G	Cessation	0.714	0.16	**0.033**	1.306	0.15	0.080	0.967	0.911
			CurSmo	0.716	0.15	**0.028**	1.216	0.14	0.155	0.937	0.805
			Obesity	1.228	0.09	**0.025**	0.930	0.11	0.516	1.077	0.595
15	rs2656069	G	AbVFat	0.011	0.02	0.653	-0.075	0.03	**0.013**	-0.030	0.482
			PhyAct	0.052	0.02	**0.037**	0.030	0.03	0.313	0.043	**0.024**
15	rs3885951	C	BMICat	0.082	0.04	**0.042**	-0.204	0.12	0.079	-0.040	0.776
			Obesity	1.155	0.14	0.316	0.380	0.37	**0.009**	0.698	0.516
15	rs578776	T	PhyAct	0.052	0.02	**0.036**	0.018	0.03	0.557	0.038	**0.047**
15	rs12441998	G	Cessation	0.667	0.20	**0.041**	1.228	0.13	0.106	0.922	0.791
			PkYrs	-0.048	0.02	**0.049**	-0.012	0.03	0.692	-0.033	0.076
15	rs11072774	T	BMI	0.024	0.02	0.315	0.059	0.03	**0.041**	0.039	**0.037**
			Cessation	0.589	0.22	**0.018**	0.870	0.18	0.426	0.733	0.109
			CurSmo	0.641	0.22	**0.041**	0.799	0.16	0.168	0.739	**0.020**
			Obesity	1.207	0.12	0.102	1.353	0.12	**0.011**	1.276	**0.003**
15	rs17487514	T	Obesity	1.021	0.10	0.837	0.657	0.20	**0.037**	0.848	0.450
15	rs16970006	C	BMI	0.018	0.02	0.471	0.066	0.03	**0.023**	0.039	0.102
			BMICat	0.040	0.05	0.418	0.218	0.10	**0.034**	0.106	0.217
			Cessation	0.435	0.40	**0.039**	1.848	0.39	0.115	0.900	0.884
			CurSmo	0.437	0.39	**0.036**	1.365	0.33	0.347	0.788	0.676
			Obesity	1.249	0.17	0.191	1.952	0.27	**0.014**	1.484	0.070
15	rs11072794	T	Cessation	0.652	0.18	**0.020**	1.364	0.15	**0.035**	0.951	0.891
			CurSmo	0.701	0.18	**0.045**	1.318	0.14	**0.044**	0.971	0.926
15	rs7177699	C	CurSmo	1.336	0.14	**0.046**	1.005	0.16	0.973	1.167	0.278
15	rs4380028	A	AbVFat	0.038	0.02	0.113	0.051	0.03	0.099	0.043	**0.023**
15	rs11072810	T	Obesity	0.986	0.09	0.875	0.826	0.09	**0.039**	0.902	0.243
16	rs802698	A	PhyAct	-0.053	0.02	**0.033**	0.017	0.03	0.563	-0.020	0.572
19	rs3889806	A	Cessation	0.674	0.17	**0.021**	0.876	0.20	0.500	0.755	**0.031**
			CurSmo	0.717	0.16	**0.039**	0.926	0.19	0.678	0.801	0.080
			PhyAct	-0.019	0.02	0.435	-0.072	0.03	**0.016**	-0.043	0.103
19	rs7251950	T	Cessation	1.386	0.16	**0.038**	1.119	0.17	0.520	1.259	**0.049**

SUPPLEMENTARY MATERIAL

Figure 34: Daily cigarette consumption by BMI and sex in the VA30k sample



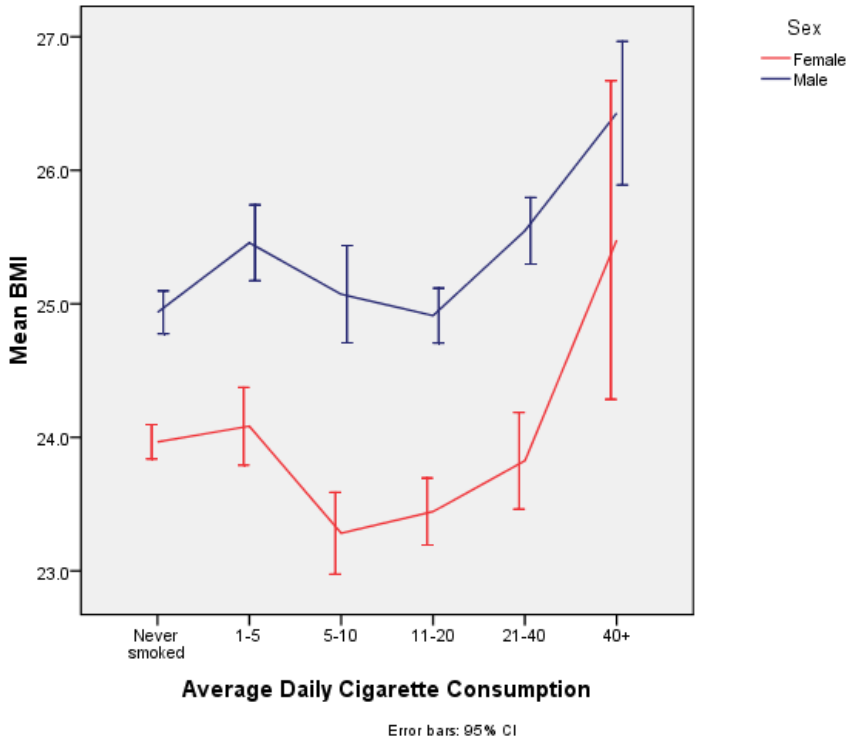
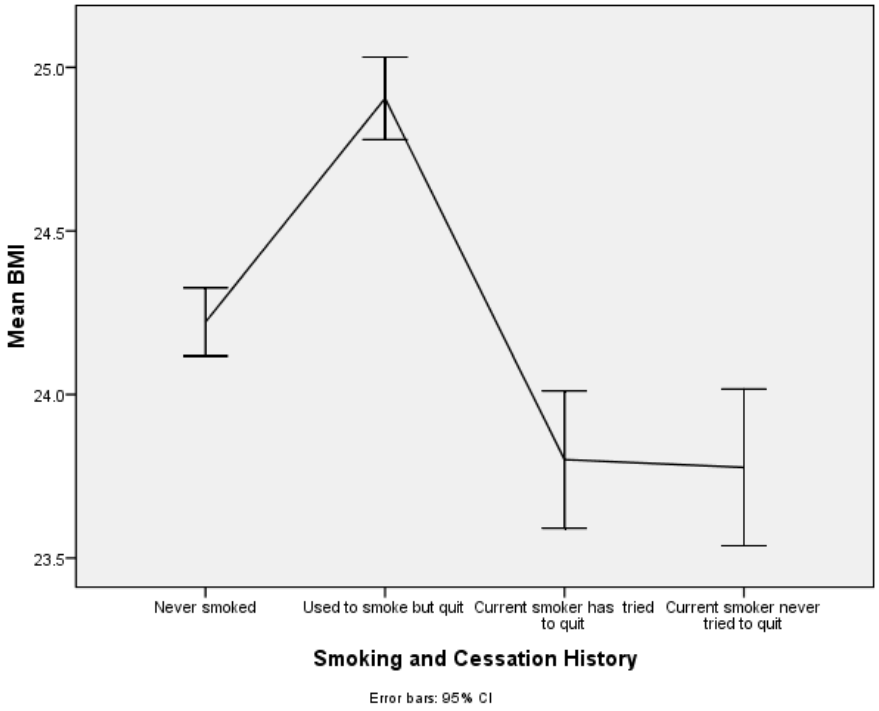


Figure 35: Smoking history by BMI and sex in the VA30k sample



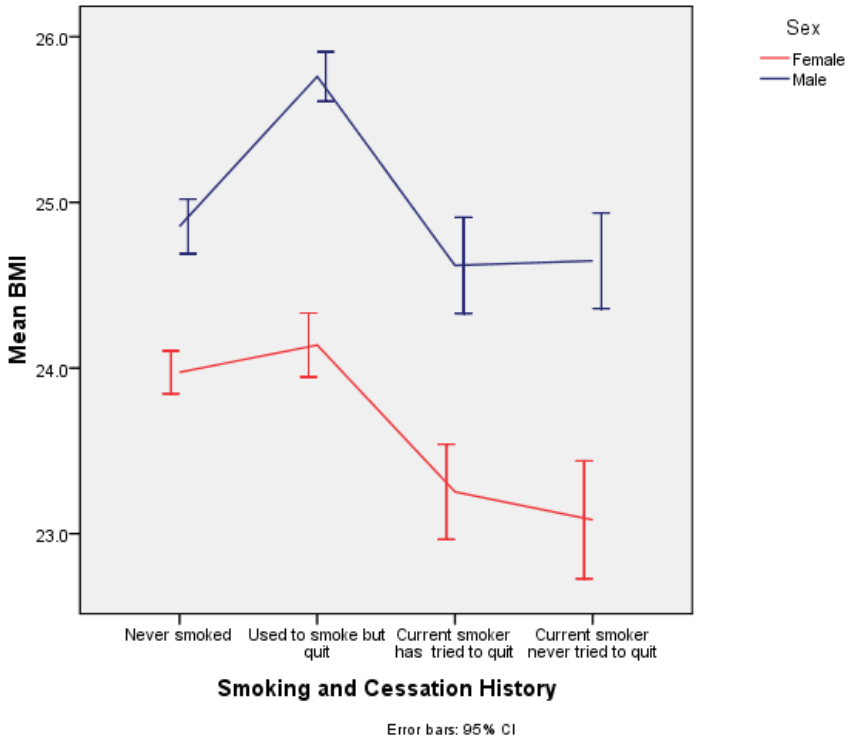


Figure 36: Partial modified CCC model path diagram for BMI, smoking initiation and nicotine dependence

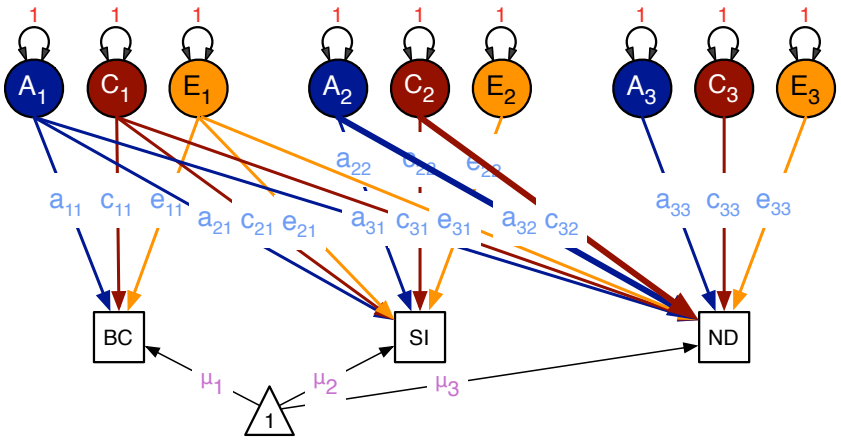


Figure 37: CCC path estimates for females (VA30k)

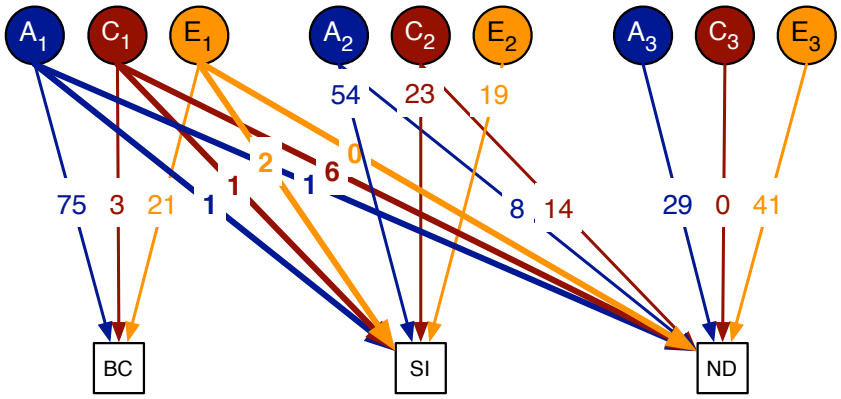
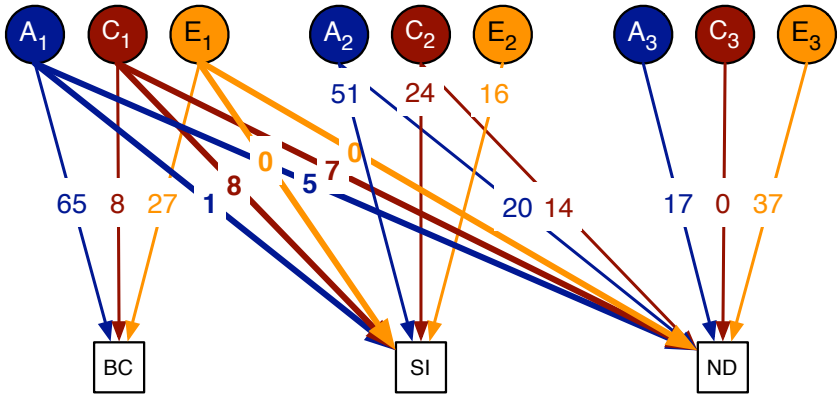


Figure 38: CCC path estimates for males (VA30k)



Chapter 9: Global Discussion

Obesity is a serious public health crisis and recent estimates of its incidence are the highest in United States history, with 35% and 17% of American adults and children affected, respectively (2). The clinical definition of adult obesity is operationalized as a body mass index (BMI) greater than 30 kg/m^2. Although the prevalence of common obesity has increased dramatically over the past 30 years–largely thought to be due to changes in the environment, such as high calorie diets and sedentary lifestyles—twin and family studies have shown consistently that relative body weight is under considerable genetic influence in both children and adults, and heritability estimates range from 40% to 90% (35, 51-54).

Given the large heritability estimates reported for BMI, molecular genetic approaches represent a useful tool with which to examine underlying mechanisms of and genetic susceptibility to obesity. To date, a number of approaches have been utilized to identify BMI/obesity-associated genes including candidate gene, linkage and association studies. While candidate gene and linkage studies have been useful in detecting genetic factors of large effect for rare forms of obesity, they have proven relatively unsuccessful for discovering genes of relatively small effect, such as those thought to underlie genetic liability to common complex obesity and BMI.

Genome-wide association studies (GWAS) have successfully identified polymorphisms influencing numerous complex traits and diseases (72). However, this approach has been met with important limitations. A number of potential factors have been proposed that may reduce the power of this methodology in general, as well as for the field of common complex obesity specifically. The survey of limitations presented in Chapter 1 highlights the following issues: replication of variants with small effects, utility of risk prediction, generalizability to multiple racial groups and across the lifespan and affects of comorbidity with other traits and disorders. The research reported herein attempts to address many of these issues, towards developing improved methods to delineate the genetics of BMI and common complex obesity, along with the corresponding associations with depression symptoms and smoking behavior. In the subsequent discussion, we first summarize key findings from each of the chapters, discuss limitations of this research and propose extensions for future research.

Research findings

This research integrated clinical, twin, and genetic association studies to further our understanding of the genetics of BMI and common complex obesity in the context of genetic risk sum scores (GRSS), clinical risk prediction, development across adolescence into adulthood, and comorbidity with depression symptoms and smoking behavior. A summary of the dissertation studies appears in Figure 39. The first three studies (Chapters 2-4) incorporated GRSS methodology, which effectively summarizes the effects of a number of risk alleles into a composite score. In Chapter 2, the MGS-C sample was used for proof-of-principle of this methodology, that is, the use of a GRSS as an alternative form of replication. The MGS-C had limited power to detect the previously BMI-associated variants individually but in aggregate, as a count score, was found to be highly

associated with BMI (p-value = $3.19x10^{-6}$) but explained a limited amount of the variance (0.66%). However, estimates of the area under receiver operator criteria curve (AUC) indicated that the GRSS and covariates significantly predicted overweight and obesity classification with maximum discriminative ability for predicting class III obesity (AUC=0.697). An additional finding was that the GRSS was associated in both European- and African-Americans, despite the fact that the BMI-associated variants were catalogued from meta-analyses of primarily European descent.

In Chapter 3, we extended this GRSS methodology by constructing scores from proxy versus imputed SNPs and count versus weighted methods. The weighted SNP-GRSS constructed from imputed probabilities of risk alleles performed best and was highly associated with BMI ($p=4.3x10^{-16}$), accounting for 3% of the phenotypic variance. In addition to BMI-validated SNPs, common and rare BMI/obesity-associated CNVs were identified from the literature and incorporated into a score in the hopes of increasing risk prediction. Of the 84 CNVs previously reported, only a 21-kilobase deletion on 16p12.3 demonstrated evidence of association with BMI (p=0.003, frequency=16.9%) in the SAGE sample, with two CNVs showing nominal association with moderate-obesity, 1p36.1 duplications (OR=3.1, p=0.009, frequency 1.2%) and 5q13.2 deletions (OR=1.5, p=0.048, frequency 7.7%). The combined model, which included covariates, SNP-GRSS, and 16p12.3 deletion, accounted for 11.5% of phenotypic variance in BMI ($p=3.34x10^{-54}$) and AUC estimates significantly predicted obesity classification with maximum discriminative ability for morbid-obesity (AUC = 0.750). These results illustrate how prediction algorithms may be improved by incorporating validated effect-sizes and allelic probabilities. Furthermore, in agreement with Chapter 2, the GRSS was associated in both European- and African-Americans despite the BMI-associated variants being catalogued from meta-analyses primarily of European descent.

Because there has been only limited research on when during development BMI-associated variants begin to influence BMI, we utilize in Chapter 4 the ABD longitudinal twin study in order to assess the effects of adult-validated BMI-SNPs across adolescence into adulthood (age 8 to 18). BMI was found to be highly heritable, accounting for 74-91% of the variance over the course of adolescent development and, furthermore, modeling indicated multiple genetic factors that contributed to BMI liability, including a genetic factor that loaded across development, a second common genetic factor that loaded later in adolescence and time-specific genetic factors important in mid-adolescence. Additionally, shared environmental effects were found to account for significant portions of the phenotypic variance (1-18%) for ages 11-16 in females and ages 8-14 in males. A unique environmental factor accounted for 2-13% of the phenotypic variance across development. To better understand the importance of adult BMI-associated genetic variants across adolescent development, we tested a weighted GRSS as an effect on latent genetic factors as well as on mean BMI. Preliminary results indicated that the GRSS was best modeled as an effect on mean BMI at each age group, suggesting association across development with the magnitude of the effect differing at each time point considered and ranged in effect from 0.05 to 2.4 kg/m^2 change in BMI. The GRSS accounted for 1-2.3% of the phenotypic variance in BMI across adolescence. To our knowledge, this is the first study of BMI to incorporate GRSS methodology in the context of variance decomposition.

In Chapters 5 through 8, BMI and common complex obesity are approached from the perspective of comorbidity through phenotypic and genetic associations with binge eating disorder (BED), depression symptoms and smoking behavior. In Chapters 5 and 6, we used the UofMN study, a clinical sample of overweight and obese women with and without BED, to examine the relationship of BED, food intake and internalizing symptoms of depression and anxiety. In Chapter 5, energy intake and energy expenditure were assessed by multiple methods to potentially identify differences in food intake, metabolism and accuracy of self-reported food intake in obese groups with and without BED. The results indicated no between group differences in total daily energy expenditure (TDEE), basal metabolic rate (BMR) or thermal effect of food (TEF). According to dietary recall data, the BED group had significantly higher caloric intake on binge eating episode days than non-binge days (3255 vs. 2343 kilocalories (kcal)). No difference was observed between BED non-binge day intake and control group intake (2233 vs. 2140 kcal). We observed similar results for food log data and laboratory measured intake. Our data suggest that increased energy intake reported by BED individuals is due to increased food consumption and, critically, not metabolic differences. When comparing TDEE to data on dietary recall and food log, both groups displayed significant underreporting of caloric intake of similar magnitudes ranging 20-33%. Furthermore, predicted energy requirements estimated via the Harris-Benedict equation underestimated measured TDEE by 23-24%. These results, taken together, provide support for under-reporting of food intake by both BED and non-BED obese groups.

In Chapter 6, we used the UofMN sample to examine models by which BED and internalizing symptoms of depression and anxiety influence food intake in overweight/obese women. The BED group was found to endorse significantly more symptoms of depression (10.1 vs. 4.8, p=0.005) and anxiety (8.5 vs. 2.7, p=0.003). Linear regression indicated that BED diagnosis and internalizing symptoms accounted for 30% of the variance in kcal-intake ($F(3,28)=4.0$, p=0.017). Results from path analysis suggested that BED mediates the influence of internalizing symptoms on total kcal-intake (empirical $p<0.001$). The associations between internalizing symptoms and food intake are best described as acting indirectly through a BED diagnosis. This suggests that symptoms of depression and anxiety influence whether an individual engages in binge eating, which itself influences kcal-intake. Improved understanding of the mechanisms underlying the associations between mood, binge eating and food intake will facilitate the development of more effective prevention and treatment strategies for both BED and obesity.

In Chapters 7 and 8, associations between BMI, depression symptoms and smoking behavior were examined by two different types of genetically informative samples: twin studies and GWAS. In Chapter 7, twin study methodology was utilized in order to investigate whether shared genetic and/or environmental liability is responsible for phenotypic associations found between BMI, depression symptoms, and impulsivity in the VA30k sample. A significant quadratic relationship was found between BMI and depression symptoms, indicating that those individuals with the highest and the lowest BMI were more likely to endorse higher depression scores. Bivariate twin modeling results did not indicate a significant genetic or environmental correlation between BMI and depression symptoms. However, significant genetic and environmental correlations

were found between BMI and impulsivity (rG =0.115, rE=0.046), as well as a significant genetic correlation between depression and impulsivity (rG=0.075). Trivariate independent pathway twin modeling indicated shared genetic and environmental liability between these traits and a common genetic factor accounting for 2-16% of the genetic variance in these traits. In females, an environmental factor common to BMI and impulsivity accounted for 0.5% of the environmental variance in BMI and 62% in impulsivity. In males, an environmental factor common to depression symptoms and impulsivity accounted for 0.5% of the environmental variance in depression symptoms and 56% in impulsivity. Our findings suggested partially shared genetic and environmental risk between BMI, depression symptoms and impulsivity.

The purpose of Chapter 8 was to investigate whether genetic variants previously identified to be associated with either BMI or smoking behavior were common to multiple behaviors or were trait-specific in the HABC study. Phenotypic associations indicated current smokers had significantly lower BMI and abdominal visceral fat than "never" or former smokers in both sexes. In total, three BMI-associated SNPs demonstrated nominally significant associations with smoking traits: rs1900273 in *STK33*, rs2145270 near *BMP2* and rs12127438 at the 1q42.2 locus. Additionally, three smoking behavior-associated SNPs were found to be nominally associated with body composition variables: rs11072774 in *CHRNB4*, rs2640732 in *SCARA3* and rs6945244 in *PDE1C*. Our preliminary findings are suggestive of partially shared genetic risk between smoking and body composition in a sample of European- and African-Americans.

Limitations and extensions

The findings reported herein are best interpreted within the context of several limitations. First, although SNP-GRSSs were significantly associated with BMI, they only accounted for a limited proportion of the phenotypic variance (0.5-3%) and, accordingly, obesity risk prediction based on these scores was not found to have clinical utility. Moreover, while it was hoped that by including an additional class of genetic variants (i.e., CNVs) we would be able to account for more of the phenotypic variance in BMI, all but three of the CNVs catalogued from the literature failed to demonstrate evidence of association with BMI or obesity, even when tested in aggregate. However, as large-scale exome and genome sequencing initiatives identify lower frequency variants and other types of variation such as INDELs, the framework we have provided for integrating common and rare variation may be applied.

There are potentially several other extensions to GRSS methodology. For example, the GRSS reported here were constructed from variants that met genome-wide significance. Alternatively, these scores could be constructed from a wider significance threshold to determine the probability level that captures maximal predictive ability. Furthermore, an important extension of an integrated model of BMI and obesity is to incorporate the moderating effects of the environment. At least two of the BMI-validated SNPs exhibit gene by environment interactions. Future research should incorporate environmental variables into models of disease and risk prediction, as consideration of only genetic effects will surely be of limited potential.

Several of the reported phenotypic associations indicated a significant quadratic association, including age and BMI. Additionally, a quadratic association was found

between BMI and depression symptoms. This finding could explain, in part, conflicting reported findings on the nature of the association of BMI and depression, and further highlights the importance of addressing the possibility of higher order associations between variables (i.e., quadratic, cubic). Furthermore, there are known limitations of structural equation modeling for the handling of nonlinear relationships. Additional research is needed to determine the effect of curvilinear relationships on variance decomposition methodology and parameter estimates.

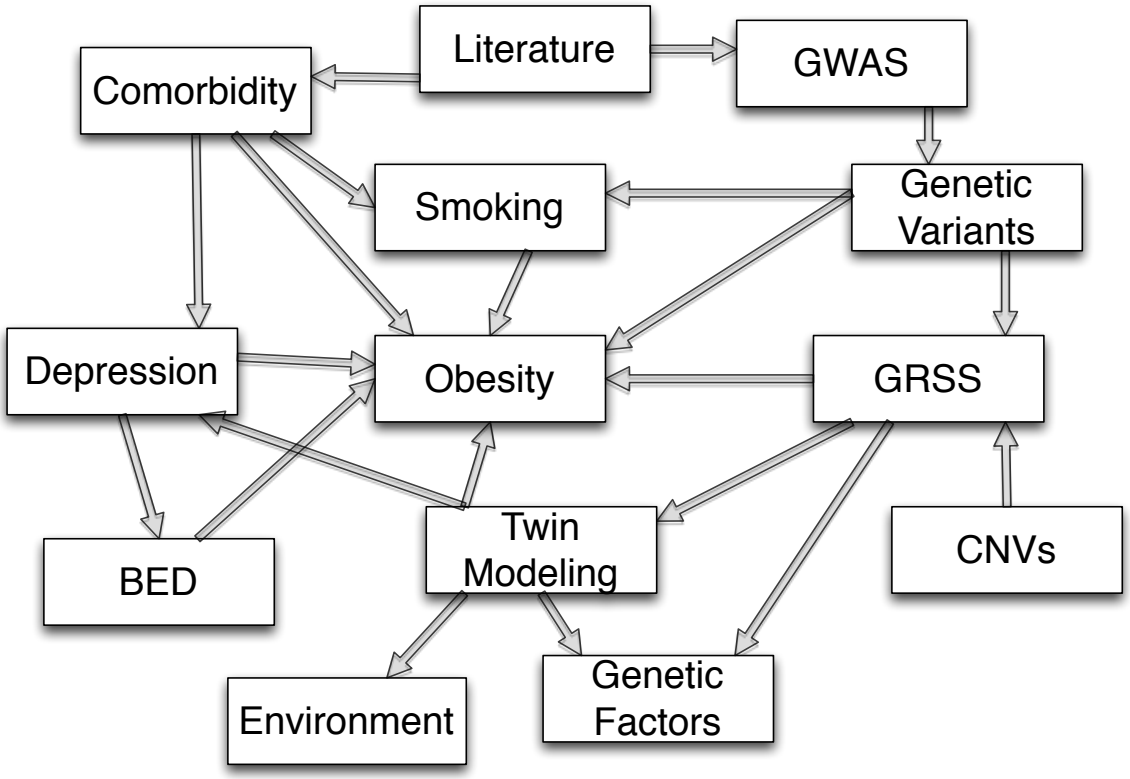
Another limitation of this research was the application of only a few of the potentially relevant latent variable twin models. Other longitudinal models, which would be particularly insightful, are simplex and growth curves, as they allow for the assessment of the contributions of variance components and genetic variants on innovations, transmissions and rate of change of BMI across time. In addition, since C and D components cannot me modeled simultaneously in classical twin designs, future models might utilize the extended twin design to determine the effect of each of these sources of variance among others such as assortative mating. Of particular interest is the application of models of comorbidity to potentially determine the direction of effect, as longitudinal phenotypic studies have found a reciprocal association between obesity and depression. Future research should apply models of comorbidity and test direction of causation in genetically informative samples. Our results suggest that there is partially shared genetic risk between BMI, depression symptoms and impulsivity and BMI and smoking behavior. More studies are needed to determine how correlated liability affects gene-finding efforts.

Closing remarks

Given the seriousness of the global obesity epidemic among both children and adults, research elucidating the genetic and environmental liability to BMI and development of obesity is essential. It is well recognized that excess body weight is the result of positive energy balance, that is, excess caloric intake relative to energy expenditure. Although energy balance appears straightforward, its relationship with obesity is quite complex and involves the interplay of genetic, environmental, and psychological determinants. Despite twin and family studies consistently demonstrating that relative body weight is under considerable genetic influence in both children and adults, only a limited number of genetic variants have been identified to date and these account for only a fraction of the heritability. The so-called "missing heritability" has been speculated to reside in lower frequency and other classes of variants yet to be elucidated by the holy grail of molecular genetic studies—whole-genome sequencing. Longitudinal twin studies indicate there are multiple genetic and environmental factors that persist across time, as well as time-specific factors, that influence relative body weight. However, most genetic association studies have been performed on cross-sectional studies ignoring the potential confounders of development. Furthermore, BMI and obesity are associated and comorbid with multiple traits and diseases, and studies have demonstrated correlated liability between traits. Nonetheless, most genetic association studies do not account for effects of correlated liability beyond the use of a few basic covariates. This next era of gene finding efforts by large-scale sequencing will certainly identify additional genetic variation and likely shed light on new pathways

involved in disease etiology. However, to fully understand common complex obesity we need to move beyond the rather simplistic model of performing linear associations between genetic variant and "trait" and move towards building integrated models incorporating development, comorbidity, and, importantly, effects of the environment.

Figure 39: Summary of dissertation studies



Reference List

1. Caballero B. The global epidemic of obesity: an overview. *Epidemiol Rev* 2007; **29**: 1-5.

2. Ogden C, Carroll M, Kit B, Flegal K. Prevalence of obesity in the United States, 2009-2010. *NCHS data brief* 2012;: 1-8.

3. Dietz WH, Bellizzi MC. Introduction: the use of body mass index to assess obesity in children. *Am J Clin Nutr* 1999; **70**: 123S-125S.

4. Lesser G. Issues in body fatness measurement. *Arch Intern Med* 2009; **169**: 636-636.

5. Gallagher D, Visser M, Sepúlveda D, Pierson RN, Harris T, Heymsfield SB. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? *Am J Epidemiol* 1996; **143**: 228-239.

6. Romero Corral A, Somers VK, Sierra Johnson J, Thomas RJ, Collazo Clavell ML, Korinek J, et al. Accuracy of body mass index in diagnosing obesity in the adult general population. *Int J Obes* 2008; **32**: 959-966.

7. Deurenberg P, Yap M. The assessment of obesity: methods for measuring body fat and global prevalence of obesity. *Bailliere's best practice research.Clinical endocrinology metabolism* 1999; **13**: 1-11.

8. Deurenberg P, Deurenberg Yap M. Validity of body composition methods across ethnic population groups. *Forum Nutr* 2003; **56**: 299-301.

9. Freedman D, Sherry B. The validity of BMI as an indicator of body fatness and risk among children. *Pediatrics* 2009; **124 Suppl 1**: S23-S34.

10. Freedman D, Wang J, Ogden C, Thornton J, Mei Z, Pierson R, et al. The prediction of body fatness by BMI and skinfold thicknesses among children and adolescents. *Ann Hum Biol* 2007; **34**: 183-194.

11. Ogden CL, Carroll MD, Flegal KM. Epidemiologic trends in overweight and obesity. *Endocrinol Metab Clin North Am* 2003; **32**: 741-60, vii.

12. Ogden CL, Yanovski SZ, Carroll MD, Flegal KM. The epidemiology of obesity. *Gastroenterology* 2007; **132**: 2087-2102.

13. Apovian C, Gokce N. Obesity and cardiovascular disease. *Circulation* 2012; **125**: 1178-1182.

14. Chen L, Magliano D, Zimmet P. The worldwide epidemiology of type 2 diabetes mellitus-present and future perspectives. *Nature reviews.Endocrinology* 2011;.

15. Temelkova Kurktschiev T, Stefanov T. Lifestyle and Genetics in Obesity and type 2 Diabetes. *Experimental and clinical endocrinology diabetes* 2011;.

16. Faulds M, Dahlman Wright K. Metabolic diseases and cancer risk. *Curr Opin Oncol* 2012; **24**: 58-61.

17. de Wit L, Fokkema M, van Straten A, Lamers F, Cuijpers P, Penninx BWJH. Depressive and anxiety disorders and the association with obesity, physical, and social activities. *Depress Anxiety* 2010; **27**: 1057-1065.

18. Carpiniello B, Pinna F, Pillai G, Nonnoi V, Pisano E, Corrias S, et al. Obesity and psychopathology. A study of psychiatric comorbidity among patients attending a specialist obesity unit. *Epidemiol Psichiatr Soc* 2009; **18**: 119-127.

19. Duncan AE, Grant JD, Bucholz KK, Madden PA, Heath AC. Relationship between body mass index, alcohol use, and alcohol misuse in a young adult female twin sample. *J Stud Alcohol Drugs* 2009; **70**: 458-466.

20. Strine TW, Mokdad AH, Dube SR, Balluz LS, Gonzalez O, Berry JT, et al. The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults. *Gen Hosp Psychiatry* 2008; **30**: 127-137.

21. Grucza RA, Przybeck TR, Cloninger CR. Prevalence and correlates of binge eating disorder in a community sample. *Compr Psychiatry* 2007; **48**: 124-131.

22. Dietz WH. Health consequences of obesity in youth: childhood predictors of adult disease. *Pediatrics* 1998; **101**: 518-525.

23. Schwartz MB, Puhl R. Childhood obesity: a societal problem to solve. *Obesity reviews* 2003; **4**: 57-71.

24. Whitlock E, Williams S, Gold R, Smith P, Shipman S. Screening and interventions for childhood overweight: a summary of evidence for the US Preventive Services Task Force. *Pediatrics* 2005; **116**: e125-e144.

25. Freedman D, Mei Z, Srinivasan S, Berenson G, Dietz W. Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa Heart Study. *J Pediatr* 2007; **150**: 12-17.e2.

26. Serdula MK, Ivery D, Coates RJ, Freedman DS, Williamson DF, Byers T. Do obese children become obese adults? A review of the literature. *Prev Med* 1993; **22**: 167-177.

27. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med* 1997; **337**: 869-873.

28. Parsons TJ, Power C, Logan S, Summerbell CD. Childhood predictors of adult obesity: a systematic review. *Int J Obes* 1999; **23 Suppl 8**: S1-107.

29. Baird J, Fisher D, Lucas P, Kleijnen J, Roberts H, Law C. Being big or growing fast: systematic review of size and growth in infancy and later obesity. *BMJ.British medical journal* 2005; **331**: 929-929.

30. Biro F, Wien M. Childhood obesity and adult morbidities. *Am J Clin Nutr* 2010; **91**: 1499S-1505S.

31. McAllister E, Dhurandhar N, Keith S, Aronne L, Barger J, Baskin M, et al. Ten putative contributors to the obesity epidemic. *Crit Rev Food Sci Nutr* 2009; **49**: 868-913.

32. Bleich S, Cutler D, Murray C, Adams A. Why is the developed world obese? *Annu Rev Public Health* 2008; **29**: 273-295.

33. Bleich SN, Ku R, Wang YC. Relative contribution of energy intake and energy expenditure to childhood obesity: a review of the literature and directions for future research. *Int J Obes* 2011; **35**: 1-15.

34. Stunkard AJ, Foch TT, Hrubec Z. A twin study of human obesity. *JAMA (Chicago, Ill.)* 1986; **256**: 51-54.

35. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 1997; **27**: 325-351.

36. Silventoinen K, Kaprio J. Genetics of Tracking of Body Mass Index from Birth to Late Middle Age: Evidence from Twin and Family Studies. *Obesity Facts* 2009; **2**: 196.

37. O'Rahilly S, Farooqi IS, Yeo GSH, Challis B. Minireview: human obesity-lessons from monogenic disorders. *Endocrinology* 2003; **144**: 3757-3764.

38. Mutch D, Clément K. Unraveling the genetics of human obesity. *PLOS Genetics* 2006; **2**: e188-e188.

39. Hinney A, Vogel CIG, Hebebrand J. From monogenic to polygenic obesity: recent advances. *European child adolescent psychiatry* 2010; **19**: 297.

40. Beales P. Obesity in single gene disorders. *Prog Mol Biol Transl Sci* 2010; **94**: 125-157.

41. Farooqi IS, Keogh J, Yeo GSH, Lank E, Cheetham T, O'Rahilly S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* 2003; **348**: 1085-1095.

42. Loos RJF. The genetic epidemiology of melanocortin 4 receptor variants. *Eur J Pharmacol* 2011; **660**: 156-164.

43. Yeo GS, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S. A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet* 1998; **20**: 111-112.

44. Vaisse C, Clement K, Guy Grand B, Froguel P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet* 1998; **20**: 113-114.

45. Farooqi IS, Wangensteen T, Collins S, Kimber W, Matarese G, Keogh J, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med* 2007; **356**: 237-247.

46. Clément K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* 1998; **392**: 398-401.

47. Hinney A, Vogel CIG, Hebebrand J. From monogenic to polygenic obesity: recent advances. *European child adolescent psychiatry* 2010; **19**: 297-310.

48. Rankinen T, Zuberi A, Chagnon Y, Weisnagel SJ, Argyropoulos G, Walts B, et al. The human obesity gene map: the 2005 update. *Obesity* 2006; **14**: 529-644.

49. Ranadive S, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am* 2008; **37**: 733-51, x.

50. Farooqi IS, O'Rahilly S. Monogenic obesity in humans. *Annu Rev Med* 2005; **56**: 443-458.

51. Schousboe K, Willemsen G, Kyvik K, Mortensen J, Boomsma D, Cornes B, et al. Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin research* 2003; **6**: 409-421.

52. Silventoinen K, Kaprio J. Genetics of tracking of body mass index from birth to late middle age: evidence from twin and family studies. *Obesity facts* 2009; **2**: 196-202.

53. Silventoinen K, Rokholm B, Kaprio J, Sørensen TIA. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. *Int J Obes* 2010; **34**: 29-40.

54. Dubois L, Ohm Kyvik K, Girard M, Tatone Tokuda F, Prusse D, Hjelmborg J, et al. Genetic and Environmental Contributions to Weight, Height, and BMI from Birth to 19 Years of Age: An International Study of Over 12,000 Twin Pairs. *PLoS ONE* 2012; **7**: e30153-e30153.

55. Silventoinen K, Rokholm B, Kaprio J, Sørensen TIA. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. *Int J Obes* 2010; **34**: 29-40.

56. Harris JR, Tambs K, Magnus P. Sex-specific effects for body mass index in the new Norwegian twin panel. *Genet Epidemiol* 1995; **12**: 251-265.

57. Bodurtha JN, Mosteller M, Hewitt JK, Nance WE, Eaves LJ, Moskowitz WB, et al. Genetic analysis of anthropometric measures in 11-year-old twins: the Medical College of Virginia Twin Study. *Pediatr Res* 1990; **28**: 1-4.

58. Ortega Alonso A, Pietiläinen K, Silventoinen K, Saarni S, Kaprio J. Genetic and environmental factors influencing BMI development from adolescence to young adulthood. *Behav Genet* 2012; **42**: 73-85.

59. Hur Y. Sex difference in heritability of BMI in South Korean adolescent twins. *Obesity* 2007; **15**: 2908-2911.

60. Cornes B, Zhu G, Martin N. Sex differences in genetic variation in weight: a longitudinal study of body mass index in adolescent twins. *Behav Genet* 2007; **37**: 648-660.

61. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature* 1994; **372**: 425-432.

62. Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, Wareham NJ, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* 1997; **387**: 903-908.

63. Speliotes E,K., Willer C,J., Berndt S,I., Monda K,L., Thorleifsson,Gudmar, Jackson A,U., et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937.

64. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**: 25-34.

65. Thorleifsson G, Walters GB, Gudbjartsson D, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41**: 18.

66. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; **40**: 768-775.

67. Loos RJF. Genetic determinants of common obesity and their value in prediction. *Best practice research.Clinical endocrinology metabolism* 2012; **26**: 211-226.

68. Saunders C, Chiodini B, Sham P, Lewis C, Abkevich V, Adeyemo A, et al. Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity* 2007; **15**: 2263-2275.

69. Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; **69**: 936-950.

70. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**: 851-861.

71. Hirschhorn J, Lettre G. Progress in genome-wide association studies of human height. *Horm Res* 2009; **71 Suppl 2**: 5.

72. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661-678.

73. Schork N, Murray S, Frazer K, Topol E. Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics development* 2009; **19**: 212-219.

74. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, et al. Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 2007; **80**: 779-791.

75. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463**: 671.

76. Bochukova E, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010; **463**: 666.

77. Wang K, Li W, Glessner J, Grant SFA, Hakonarson H, Price RA. Large copy-number variations are enriched in cases with moderate to extreme obesity. *Diabetes* 2010; **59**: 2690-2694.

78. Bachmann Gagescu R, Mefford H, Cowan C, Glew G, Hing A, Wallace S, et al. Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine* 2010; **12**: 641.

79. Glessner J, Bradfield J, Wang K, Takahashi N, Zhang H, Sleiman P, et al. A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *Am J Hum Genet* 2010; **87**: 661.

80. Shinawi M, Sahoo T, Maranda B, Skinner SA, Skinner C, Chinault C, et al. 11p14.1 microdeletions associated with ADHD, autism, developmental delay, and obesity. *American journal of medical genetics.Part A* 2011; **155A**: 1272-1280.

81. Jacquemont,Sebastien, Reymond,Alexandre, Zufferey,Flore, Harewood,Louise, Walters R,G., Kutalik,Zoltan, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**: 97-102.

82. Sofos E, Pescosolido M, Quintos J, Abuelo D, Gunn S, Hovanes K, et al. A novel familial 11p15.4 microduplication associated with intellectual disability, dysmorphic features, and obesity with involvement of the ZNF214 gene. *American journal of medical genetics.Part A* 2011;.

83. Moran N. UK10K Seeks to Sequence 10,000 Individual Genomes. *BioWorld* 2010;: 1-3.

84. Mather A, Cox B, Enns M, Sareen J. Associations of obesity with psychiatric disorders and suicidal behaviors in a nationally representative sample. *J Psychosom Res* 2009; **66**: 277-285.

85. Berkowitz R, Fabricatore A. Obesity, psychiatric status, and psychiatric medications. *Psychiatr Clin North Am* 2011; **34**: 747-764.

86. Petry N, Barry D, Pietrzak R, Wagner J. Overweight and obesity are associated with psychiatric disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychosom Med* 2008; **70**: 288.

87. American Psychiatric Association. (ed) *Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC:* , 1994.

88. Spitzer RL, Yanovski S, Wadden T, Wing R, Marcus MD, Stunkard A, et al. Binge eating disorder: its further validation in a multisite study. *Int J Eat Disord* 1993; **13**: 137-153.

89. de Zwaan M. Binge eating disorder and obesity. *Int J Obes Relat Metab Disord* 2001; **25 Suppl 1**: S51-5.

90. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; **62**: 617-627.

91. Blaine B. Does Depression Cause Obesity?: A Meta-analysis of Longitudinal Studies of Depression and Weight Control. *J Health Psychol* 2008; **13**: 1190-1197.

92. Luppino F, de Wit L, Bouvy P, Stijnen T, Cuijpers P, Penninx BWJH, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry* 2010; **67**: 220-229.

93. Wing, Matthews, Kuller, Meilahn, Plantinga,. Waist to hip ratio in middle-aged women. Associations with behavioral and psychosocial factors and with changes in cardiovascular risk factors. *Arteriosclerosis and thrombosis* 1991; **11**: 1250.

94. Istvan, Zavela, Weidner,. Body weight and psychological distress in NHANES I. *Int J Obes* 1992; **16**: 999.

95. Becker, Margraf, Trke, Soeder, Neumer,. Obesity and mental illness in a representative sample of young women. *Int J Obes* 2001; **25 Suppl 1**: S5.

96. Richardson L, Garrison M, Drangsholt M, Mancl L, LeResche L. Associations between depressive symptoms and obesity during puberty. *Gen Hosp Psychiatry* 2006; **28**: 313.

97. Baumeister, Hrter,. Mental disorders in patients with obesity in comparison with healthy probands. *Int J Obes* 2007; **31**: 1155.

98. Crisp, McGuiness,. Jolly fat: relation between obesity and psychoneurosis in general population. *Br Med J* 1976; **1**: 7.

99. Stewart, Brook,. Effects of being overweight. *Am J Public Health* 1983; **73**: 171.

100. Han, Tijhuis, Lean, Seidell,. Quality of life in relation to overweight and body fat distribution. *Am J Public Health* 1998; **88**: 1814.

101. John U, Meyer C, Rumpf H, Hapke U. Relationships of psychiatric disorders with overweight and obesity in an adult general population. *Obes Res* 2005; **13**: 101.

102. Kress A, Peterson M, Hartzell M. Association between obesity and depressive symptoms among U.S. Military active duty service personnel, 2002. *J Psychosom Res* 2006; **60**: 263.

103. de Wit LM, van Straten A, van Herten M, Penninx BW, Cuijpers P. Depression and body mass index, a u-shaped association. *BMC Public Health* 2009; **9**: 14.

104. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119-1127.

105. Etter JF, Duc TV, Perneger TV. Validity of the Fagerström test for nicotine dependence and of the Heaviness of Smoking Index among relatively light smokers. *Addiction* 1999; **94**: 269-281.

106. Anonymous. Cigarette smoking among adults--United States, 2007. *MMWR.Morbidity and Mortality Weekly Report* 2008; **57**: 1221.

107. Anonymous. Vital signs: current cigarette smoking among adults aged >= 18 years--United States, 2005-2010. *Morb Mortal Weekly Rep* 2011; **60**: 1207-1212.

108. Centers for Disease Control and Prevention (US), National Center for Chronic Disease Prevention and Health Promotion (US), Office on Smoking and Health (US). 2010;.

109. Shimokata, Muller, Andres,. Studies in the distribution of body fat. III. Effects of cigarette smoking. *JAMA* 1989; **261**: 1169.

110. Flegal, Troiano, Pamuk, Kuczmarski, Campbell,. The influence of smoking cessation on the prevalence of overweight in the United States. *New England Journal of Medicine, The* 1995; **333**: 1165.

111. Huot, Paradis, Ledoux,. Factors associated with overweight and obesity in Quebec adults. *Int J Obes* 2004; **28**: 766.

112. Hofstetter A, Schutz Y, Jéquier E, Wahren J. Increased 24-hour energy expenditure in cigarette smokers. *N Engl J Med* 1986; **314**: 79-82.

113. Chiolero A, Faeh D, Paccaud F, Cornuz J. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *Am J Clin Nutr* 2008; **87**: 801.

114. Chen H, Saad S, Sandow S, Bertrand P. Cigarette smoking and brain regulation of energy homeostasis. *Frontiers in Pharmacology* 2012; **3**: 147-147.

115. Perkins KA, Epstein LH, Marks BL, Stiller RL, Jacob RG. The effect of nicotine on energy expenditure during light physical activity. *N Engl J Med* 1989; **320**: 898-903.

116. Perkins KA. Metabolic effects of cigarette smoking. *J Appl Physiol* 1992; **72**: 401-409.

117. Filozof C, Fernández Pinilla MC, Fernández Cruz A. Smoking cessation and weight gain. *Obesity reviews* 2004; **5**: 95-103.

118. Aubin H, Farley A, Lycett D, Lahmek P, Aveyard P. Weight gain in smokers after quitting cigarettes: meta-analysis. *BMJ.British medical journal* 2012; **345**: e4439-e4439.

119. Bamia, Trichopoulou, Lenas, Trichopoulos,. Tobacco smoking in relation to body fat mass and distribution in a general population sample. *Int J Obes* 2004; **28**: 1091.

120. John, Hanke, Rumpf H, Thyrian,. Smoking status, cigarettes per day, and their relationship to overweight and obesity among former and current smokers in a national adult general population sample. *Int J Obes* 2005; **29**: 1289.

121. Chiolero A, Jacot-Sadowski I, Faeh D, Paccaud F, Cornuz J. Association of cigarettes smoked daily with obesity in a general adult population. *Obesity* 2007; **15**: 1311.

122. Barrett Connor E, Khaw KT. Cigarette smoking and increased central adiposity. *Ann Intern Med* 1989; **111**: 783-787.

123. Canoy D, Wareham N, Luben R, Welch A, Bingham S, Day N, et al. Cigarette smoking and fat distribution in 21,828 British men and women: a population-based study. *Obes Res* 2005; **13**: 1466-1475.

124. Clair C, Chiolero A, Faeh D, Cornuz J, Marques Vidal P, Paccaud F, et al. Dose-dependent positive association between cigarette smoking, abdominal obesity and body fat: cross-sectional data from a population-based survey. *BMC Public Health* 2011; **11**: 23-23.

125. Sebastiani P, Timofeev N, Dworkis D, Perls T, Steinberg M. Genome-wide association studies and the genetic dissection of complex traits. *Am J Hematol* 2009; **84**: 504-515.

126. Anonymous. Prevalence and Trends in Obesity Among US Adults, 1999-2008, January 20, 2010, Flegal et al. 303 (3): 235 — JAMA. **2012**.

127. National Center for Health Statistics. NHANES data sets and related documentation.

128. Kopelman PG. Obesity as a medical problem. *Nature* 2000; **404**: 635.

129. Ogden C, Carroll M, Curtin L, McDowell M, Tabak C, Flegal K. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA* 2006; **295**: 1549.

130. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**: 889-894.

131. Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 2007; **39**: 724-726.

132. Herbert A, Gerry N, McQueen M, Heid I, Pfeufer A, Illig T, et al. A common genetic variant is associated with adult and childhood obesity. *Science* 2006; **312**: 279.

133. Scuteri A, Sanna S, Chen W, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genetics* 2007; **3**: e115.

134. Hinney A, Nguyen T, Scherag A, Friedel S, Brnner G, Mller T, et al. Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PloS one* 2007; **2**: e1361.

135. Chambers J, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* 2008; **40**: 716.

136. Yang Q, Khoury M, Friedman J, Little J, Flanders WD. How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol* 2005; **34**: 1129-1137.

137. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008; **456**: 18-21.

138. Meigs J, Shrader P, Sullivan L, McAteer J, Fox C, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *New England Journal of Medicine, The* 2008; **359**: 2208.

139. Talmud P, Hingorani A, Cooper J, Marmot M, Brunner E, Kumari M, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ.British medical journal (Clinical research ed.)* 2010; **340**: b4838.

140. Renström F, Payne F, Nordström A, Brito E, Rolandsson O, Hallmans G, et al. Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Hum Mol Genet* 2009; **18**: 1489.

141. Zhao J, Bradfield J, Li M, Wang K, Zhang H, Kim C, et al. The role of obesity-associated loci identified in genome-wide association studies in the determination of pediatric BMI. *Obesity* 2009; **17**: 2254.

142. Li S, Zhao J, Luan J, Luben R, Rodwell S, Khaw K, et al. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am J Clin Nutr* 2010; **91**: 184.

143. Cheung CY, Tso AW, Cheung BM, Xu A, Ong KL, Fong CH, et al. Obesity Susceptibility Genetic Variants Identified from Recent Genome-Wide Association Studies: Implications in a Chinese Population. *J Clin Endocrinol Metab* 2010;.

144. Sanders A, Duan J, Levinson D, Shi J, He D, Hou C, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry* 2008; **165**: 497-506.

145. Shi J, Levinson D, Duan J, Sanders A, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**: 753-757.

146. Sanders A, Levinson D, Duan J, Dennis JM, Li R, Kendler K, et al. The Internet-based MGS2 control sample: self report of mental illness. *Am J Psychiatry* 2010; **167**: 854-865.

147. Johnson A, Handsaker R, Pulit S, Nizzari M, O'Donnell C, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938-2939.

148. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263.

149. Barrett J. Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harbor protocols* 2009; **2009**: pdb.ip71.

150. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559.

151. Evans D, Visscher P, Wray N. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009; **18**: 3525.

152. Purcell S, Wray N, Stone J, Visscher P, O'Donovan M, Sullivan P, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748.

153. Armitage P, Colton T. *Encyclopedia of biostatistics,* John Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2005.

154. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29-36.

155. Janssens ACJW, Aulchenko Y, Elefante S, Borsboom GJJM, Steyerberg E, van Duijn C. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genetics in medicine* 2006; **8**: 395-400.

156. Wang Y, Beydoun M, Liang L, Caballero B, Kumanyika S. Will all Americans become overweight or obese? estimating the progression and cost of the US obesity epidemic. *Obesity* 2008; **16**: 2323-2330.

157. Yang J, Benyamin B, McEvoy B, Gordon S, Henders A, Nyholt D, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;.

158. Flint J, Mackay TFC. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* 2009; **19**: 723-733.

159. Chambers J, Swanson V. A health assessment tool for multiple risk factors for obesity: age and sex differences in the prediction of body mass index. *Br J Nutr* 2010; **104**: 298-307.

160. Newby PK, Weismayer C, Akesson A, Tucker K, Wolk A. Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *J Nutr* 2006; **136**: 2580-2587.

161. French SA, Jeffery RW, Forster JL, McGovern PG, Kelder SH, Baxter JE. Predictors of weight change over two years among a population of working adults: the Healthy Worker Project. *Int J Obes* 1994; **18**: 145-154.

162. Jebb SA, Moore MS. Contribution of a sedentary lifestyle and inactivity to the etiology of overweight and obesity: current evidence and research issues. *Med Sci Sports Exerc* 1999; **31**: S534-S541.

163. Rampersaud E, Mitchell B, Pollin T, Fu M, Shen H, O'Connell J, et al. Physical activity and the association of common FTO gene variants with body mass index and obesity. *Arch Intern Med* 2008; **168**: 1791.

164. Lappalainen T, Tolppanen A, Kolehmainen M, Schwab U, Lindstrom J, Tuomilehto J, et al. The common variant in the FTO gene did not modify the effect of lifestyle changes on body weight: the Finnish Diabetes Prevention Study. *Obesity* 2009; **17**: 832.

165. Qi L, Kraft P, Hunter D, Hu F. The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Hum Mol Genet* 2008; **17**: 3502.

166. Brandsttter A, Lingenhel A, Zwiauer K, Strobl W, Kronenberg F. Decrease of Lp(a) during weight reduction in obese children is modified by the apo(a) kringle-IV copy number variation. *Int J Obes* 2009; **33**: 1136.

167. Razquin C, Martinez JA, Martinez-Gonzalez MA, Bes-Rastrollo M, Fernndez-Crehuet J, Marti A. A 3-year intervention with a Mediterranean diet modified the association between the rs9939609 gene variant in FTO and body weight changes. *Int J Obes* 2010; **34**: 266.

168. Fan W, Boston BA, Kesterson RA, Hruby VJ, Cone RD. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* 1997; **385**: 165-168.

169. Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, Berkemeier LR, et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 1997; **88**: 131-141.

170. Marsh DJ, Hollopeter G, Huszar D, Laufer R, Yagaloff KA, Fisher SL, et al. Response of melanocortin-4 receptor-deficient mice to anorectic and orexigenic peptides. *Nat Genet* 1999; **21**: 119-122.

171. Chen AS, Metzger JM, Trumbauer ME, Guan XM, Yu H, Frazier EG, et al. Role of the melanocortin-4 receptor in metabolic rate and food intake in mice. *Transgenic Res* 2000; **9**: 145-154.

172. Samama P, Rumennik L, Grippo J. The melanocortin receptor MCR4 controls fat consumption. *Regul Pept* 2003; **113**: 85-88.

173. Branson R, Potoczna N, Kral J, Lentes K, Hoehe M, Horber F. Binge eating as a major phenotype of melanocortin 4 receptor gene mutations. *New England Journal of Medicine, The* 2003; **348**: 1096-1103.

174. Potoczna N, Branson R, Kral J, Piec G, Steffen R, Ricklin T, et al. Gene variants and binge eating as predictors of comorbidity and outcome of treatment in severe obesity. *Journal of gastrointestinal surgery* 2004; **8**: 971-81.

175. Valladares M, Domnguez-Vsquez P, Obregn AM, Weisstaub G, Burrows R, Maiz A, et al. Melanocortin-4 receptor gene variants in Chilean families: association with childhood obesity and eating behavior. *Nutr Neurosci* 2010; **13**: 71-78.

176. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA* 2006; **295**: 1549-1555.

177. Loos RJF. Recent progress in the genetics of common obesity. *Br J Clin Pharmacol* 2009; **68**: 811.

178. Day F, Loos RJF. Developments in obesity genetics in the era of genome-wide association studies. *Journal of nutrigenetics and nutrigenomics* 2011; **4**: 222-238.

179. Yang J, Manolio T, Pasquale L, Boerwinkle E, Caporaso N, Cunningham J, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011;.

180. Peterson R, Maes H, Holmans P, Sanders A, Levinson D, Shi J, et al. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet* 2011; **129**: 221-30.

181. Li B, Leal S. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**: 311-321.

182. Bansal V, Libiger O, Torkamani A, Schork N. Statistical analysis strategies for association studies involving rare variants. *Nature reviews.Genetics* 2010; **11**: 773-785.

183. Bierut L, Agrawal A, Bucholz K, Doheny K, Laurie C, Pugh E, et al. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A* 2010; **107**: 5082-5087.

184. Cornelis M, Agrawal A, Cole J, Hansel N, Barnes K, Beaty T, et al. The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* 2010; **34**: 364-372.

185. Bierut L, Saccone N, Rice J, Goate A, Foroud T, Edenberg H, et al. Defining alcohol-related phenotypes in humans. The Collaborative Study on the Genetics of Alcoholism. *Alcohol Research and Health* 2002; **26**: 208-213.

186. Edenberg H, Bierut L, Boyce P, Cao M, Cawley S, Chiles R, et al. Description of the data from the Collaborative Study on the Genetics of Alcoholism (COGA) and single-nucleotide polymorphism genotyping for Genetic Analysis Workshop 14. *BMC genetics* 2005; **6 Suppl 1**: S2-S2.

187. Bierut L, Madden PAF, Breslau N, Johnson E, Hatsukami D, Pomerleau O, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24-35.

188. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 1994; **55**: 149-158.

189. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction* 1999; **94**: 1361-1370.

190. American Psychiatric Association., American Psychiatric Association. Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders : DSM-IV-TR*. American Psychiatric Association: Washington, DC, 2000.

191. Gauderman WJ MJ. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006;.

192. Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904.

193. Patterson N, Price A, Reich D. Population structure and eigenanalysis. *PLOS Genetics* 2006; **2**: e190.

194. Shriner D. Investigating population stratification and admixture using eigenanalysis of dense genotypes. *Heredity* 2011; **107**: 413-420.

195. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007; **17**: 1665-1674.

196. Colella S, Yau C, Taylor J, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007; **35**: 2013-2025.

197. Diskin S, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res* 2008; **36**: e126-e126.

198. Lin P, Hartz S, Wang J, Krueger R, Foroud T, Edenberg H, et al. Copy number variation accuracy in genome-wide association studies. *Hum Hered* 2011; **71**: 141-147.

199. Sanders S, Ercan Sencicek AG, Hus V, Luo R, Murtha M, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; **70**: 863-885.

200. Need A, Ge D, Weale M, Maia J, Feng S, Heinzen E, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLOS Genetics* 2009; **5**: e1000373-e1000373.

201. Howie B, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genetics* 2009; **5**: e1000529-e1000529.

202. Howie B, Marchini J, Stephens M. Genotype Imputation with Thousands of Genomes. *G3: Genes, Genomes, Genetics* 2011; **1**: 457-470.

203. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061-1073.

204. Johnson A, Handsaker R, Pulit S, Nizzari M, O'Donnell C, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938.

205. Sha B, Yang T, Zhao L, Chen X, Guo Y, Chen Y, et al. Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J Hum Genet* 2009; **54**: 199.

206. Jeon J, Shim S, Nam H, Ryu G, Hong E, Kim H, et al. Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* 2010; **11**: 426.

207. Chen Y, Liu Y, Pei Y, Yang T, Deng F, Liu X, et al. Copy Number Variations at the Prader-Willi Syndrome Region on Chromosome 15 and associations with Obesity in Whites. *Obesity* 2011;.

208. Jarick I, Vogel CIG, Scherag S, Schfer H, Hebebrand J, Hinney A, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet* 2011; **20**: 840.

209. Huang L, Teng D, Wang H, Sheng G, Liu T. Association of copy number variation in the AHI1 gene with risk of obesity in Chinese population. *European journal of endocrinology* 2012;.

210. R Development Core Team. R: A language and environment for statistical computing. 2011;.

211. Janssens ACJW, van Duijn C. Towards predictive genetic testing of complex diseases. *Eur J Epidemiol* 2006; **21**: 869-870.

212. Vergara I, Norambuena T, Ferrada E, Slater A, Melo F. StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics* 2008; **9**: 265-265.

213. McCarroll S, Altshuler D. Copy-number variation and association studies of human disease. *Nat Genet* 2007; **39**: S37-S42.

214. Cooper G, Zerr T, Kidd J, Eichler E, Nickerson D. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* 2008; **40**: 1199-1203.

215. Baross A, Delaney A, Li HI, Nayar T, Flibotte S, Qian H, et al. Assessment of algorithms for high throughput detection of genomic copy number variation in oligonucleotide microarray data. *BMC Bioinformatics* 2007; **8**: 368-368.

216. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Briefings in functional genomics and proteomics* 2009; **8**: 353-366.

217. Dellinger A, Saw S, Goh L, Seielstad M, Young T, Li Y. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res* 2010; **38**: e105-e105.

218. Zhang D, Qian Y, Akula N, Alliey Rodriguez N, Tang J, Gershon E, et al. Accuracy of CNV Detection from GWAS Data. *PLoS ONE* 2011; **6**: e14511-e14511.

219. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 2011; **29**: 512-520.

220. Eckel Passow J, Atkinson E, Maharjan S, Kardia SLR, de Andrade M. Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. *BMC Bioinformatics* 2011; **12**: 220-220.

221. Koike A, Nishida N, Yamashita D, Tokunaga K. Comparative analysis of copy number variation detection methods and database construction. *BMC genetics* 2011; **12**: 29-29.

222. Loureno S, Oliveira A, Lopes C. The effect of current and lifetime alcohol consumption on overall and central obesity. *Eur J Clin Nutr* 2012;.

223. Valladares M, Domnguez-Vsquez P, Obregn AM, Weisstaub G, Burrows R, Maiz A, et al. Melanocortin-4 receptor gene variants in Chilean families: association with childhood obesity and eating behavior. *Nutr Neurosci* 2010; **13**: 71-78.

224. Kilpeläinen, T. O. et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Medicine* 2011; **Nov;8(11):e1001116.**

225. Anonymous. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health. *Obes Res* 1998; **6 Suppl 2**: 51S-209S.

226. Wang ZQ, Ouyang Z, Wang DM, Tang XL. Heritability of blood pressure in 7- to 12-year-old Chinese twins, with special reference to body size effects. *Genet Epidemiol* 1990; **7**: 447-452.

227. Allison DB, Kaprio J, Korkeila M, Koskenvuo M, Neale MC, Hayakawa K. The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int J Obes* 1996; **20**: 501-506.

228. Beunen G, Maes HH, Vlietinck R, Malina RM, Thomis M, Feys E, et al. Univariate and multivariate genetic analysis of subcutaneous fatness and fat distribution in early adolescence. *Behav Genet* 1998; **28**: 279-288.

229. Faith MS, Pietrobelli A, Nuñez C, Heo M, Heymsfield SB, Allison DB. Evidence for independent genetic influences on fat mass and body mass index in a pediatric twin sample. *Pediatrics* 1999; **104**: 61-67.

230. Pietiläinen KH, Kaprio J, Rissanen A, Winter T, Rimpelä A, Viken RJ, et al. Distribution and heritability of BMI in Finnish adolescents aged 16y and 17y: a study of 4884 twins and 2509 singletons. *Int J Obes* 1999; **23**: 107-115.

231. Silventoinen K, Pietiläinen KH, Tynelius P, Sørensen TIA, Kaprio J, Rasmussen F. Genetic and environmental factors in relative weight from birth to age 18: the Swedish young male twins study. *Int J Obes* 2007; **31**: 615-621.

232. Silventoinen K, Bartels M, Posthuma D, Estourgie-van Burk GF, Willemsen G, van Beijsterveldt,Toos C E M., et al. Genetic regulation of growth in height and weight from 3 to 12 years of age: a longitudinal study of Dutch twin children. *Twin research and human genetics* 2007; **10**: 354-363.

233. Hur Y, Kaprio J, Iacono WG, Boomsma DI, McGue M, Silventoinen K, et al. Genetic influences on the difference in variability of height, weight and body mass index between Caucasian and East Asian adolescent twins. *Int J Obes* 2008; **32**: 1455-1467.

234. Haworth CMA, Carnell S, Meaburn E, Davis OSP, Plomin R, Wardle J. Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity* 2008; **16**: 2663-2668.

235. Haworth CMA, Plomin R, Carnell S, Wardle J. Childhood obesity: genetic and environmental overlap with normal-range BMI. *Obesity* 2008; **16**: 1585-1590.

236. Wardle J, Carnell S, Haworth C, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr* 2008; **87**: 398-404.

237. Duncan A, Agrawal A, Grant J, Bucholz K, Madden PAF, Heath A. Genetic and environmental contributions to BMI in adolescent and young adult women. *Obesity* 2009; **17**: 1040-1043.

238. Lajunen H, Kaprio J, Keski Rahkonen A, Rose RJ, Pulkkinen L, Rissanen A, et al. Genetic and environmental effects on body mass index during adolescence: a prospective study among Finnish twins. *Int J Obes* 2009; **33**: 559-567.

239. Haberstick B, Lessem J, McQueen M, Boardman J, Hopfer C, Smolen A, et al. Stable genes and changing environments: body mass index across adolescence and young adulthood. *Behav Genet* 2010; **40**: 495-504.

240. Dellava J, Lichtenstein P, Kendler K. Genetic variance of body mass index from childhood to early adulthood. *Behav Genet* 2012; **42**: 86-95.

241. Nan C, Guo B, Warner C, Fowler T, Barrett T, Boomsma D, et al. Heritability of body mass index in pre-adolescence, young adulthood and late adulthood. *Eur J Epidemiol* 2012; **27**: 247-253.

242. McArdle JJ, Goldsmith HH. Alternative common factor models for multivariate biometric analyses. *Behav Genet* 1990; **20**: 569-608.

243. Loehlin J. The Cholesky approach: A cautionary note. *Behavior Genetics* 1996; **26**: 65-69.

244. McArdle JJ. Latent variable growth within behavior genetic models. *Behav Genet* 1986; **16**: 163-200.

245. Eaves LJ, Long J, Heath AC. A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav Genet* 1986; **16**: 143-162.

246. Zhao J, Bradfield J, Zhang H, Sleiman P, Kim C, Glessner J, et al. Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans. *Obesity* 2011; **19**: 2436-2439.

247. Bradfield J, Taal HR, Timpson N, Scherag A, Lecoeur C, Warrington N, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 2012;.

248. Meyer JM, Silberg JL, Simonoff E, Kendler KS, Hewitt JK. The Virginia Twin-Family Study of Adolescent Behavioral Development: assessing sample biases in demographic correlates of psychopathology. *Psychol Med* 1996; **26**: 1119-1133.

249. Hewitt JK, Silberg JL, Rutter M, Simonoff E, Meyer JM, Maes H, et al. Genetics and developmental psychopathology: 1. Phenotypic assessment in the Virginia Twin Study of Adolescent Behavioral Development. *J Child Psychol Psychiatry Allied Disciplines* 1997; **38**: 943-963.

250. Eaves LJ, Silberg JL, Meyer JM, Maes HH, Simonoff E, Pickles A, et al. Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *J Child Psychol Psychiatry Allied Disciplines* 1997; **38**: 965-980.

251. Altshuler D, Gibbs R, Peltonen L, Dermitzakis E, Schaffner S, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52-58.

252. Eaves LJ, Last KA, Young PA, Martin NG. Model-fitting approaches to the analysis of human behaviour. *Heredity* 1978; **41**: 249-320.

253. Neale MC. *Methodology for genetic studies of twins and families,* , 1992.

254. Silventoinen K, Kaprio J. Genetics of tracking of body mass index from birth to late middle age: evidence from twin and family studies. *Obesity facts* 2009; **2**: 196-202.

255. Dellava J, Lichtenstein P, Kendler K. Genetic variance of body mass index from childhood to early adulthood. *Behav Genet* 2012; **42**: 86-95.

256. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* 2011; **76**: 306-317.

257. Visscher P, Duffy D. The value of relatives with phenotypes but missing genotypes in association studies for quantitative traits. *Genet Epidemiol* 2006; **30**: 30-36.

258. Maes H, Neale M, Chen X, Chen J, Prescott C, Kendler K. A twin association study of nicotine dependence with markers in the CHRNA3 and CHRNA5 genes. *Behav Genet* 2011; **41**: 680-690.

259. Eaves L, Silberg J, Foley D, Bulik C, Maes H, Erkanli A, et al. Genetic and environmental influences on the relative timing of pubertal change. *Twin research* 2004; **7**: 471-481.

260. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. *Int J Obes* 2009; **33**: 29-36.

261. Silventoinen K, Hasselbalch A, Lallukka T, Bogl L, Pietiläinen K, Heitmann B, et al. Modification effects of physical activity and protein intake on heritability of body size and composition. *Am J Clin Nutr* 2009; **90**: 1096-1103.

262. Naukkarinen J, Rissanen A, Kaprio J, Pietilinen KH. Causes and consequences of obesity: the contribution of recent twin studies. *Int J Obes* 2011;.

263. Shin S, Miller D. A longitudinal examination of childhood maltreatment and adolescent obesity: results from the National Longitudinal Study of Adolescent Health (AddHealth) Study. *Child Abuse Neglect* 2012; **36**: 84-94.

264. Sarlio-Lähteenkorva S, Silventoinen K, Lahti Koski M, Laatikainen T, Jousilahti P. Socio-economic status and abdominal obesity among Finnish adults from 1992 to 2002. *Int J Obes* 2006; **30**: 1653-1660.

265. Midei AJ, Matthews KA. Interpersonal violence in childhood as a risk factor for obesity: a systematic review of the literature and proposed pathways. *Obesity reviews* 2011; **12**: e159-e172.

266. Lane S, Bluestone C, Burke C. Trajectories of BMI from early childhood through early adolescence: SES and psychosocial predictors. *British Journal of Health Psychology* 2012;.

267. Hornor G. Child sexual abuse: consequences and implications. *Journal of pediatric health care* 2010; **24**: 358-364.

268. Hu G, Tuomilehto J, Silventoinen K, Barengo NC, Peltonen M, Jousilahti P. The effects of physical activity and body mass index on cardiovascular, cancer and all-cause mortality among 47 212 middle-aged Finnish men and women. *Int J Obes* 2005; **29**: 894-902.

269. Benjamin A, Suchindran S, Pearce K, Rowell J, Lien L, Guyton J, et al. Gene by sex interaction for measures of obesity in the framingham heart study. *Journal of Obesity* 2011; **2011**: 329038-329038.

270. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 2010; **42**: 949-960.

271. Zhao J, Bradfield J, Li M, Wang K, Zhang H, Kim C, et al. The role of obesity-associated loci identified in genome-wide association studies in the determination of pediatric BMI. *Obesity* 2009; **17**: 2254-2257.

272. Elks C, Loos RJF, Sharp S, Langenberg C, Ring S, Timpson N, et al. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Medicine* 2010; **7**: e1000284-e1000284.

273. Elks C, Loos RJF, Hardy R, Wills A, Wong A, Wareham N, et al. Adult obesity susceptibility variants are associated with greater childhood weight gain and a faster tempo of growth: the 1946 British Birth Cohort Study. *Am J Clin Nutr* 2012; **95**: 1150-1156.

274. den Hoed M, Ekelund U, Brage S, Grontved A, Zhao J, Sharp S, et al. Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes* 2010; **59**: 2980-2988.

275. Mei H, Chen W, Jiang F, He J, Srinivasan S, Smith E, et al. Longitudinal Replication Studies of GWAS Risk SNPs Influencing Body Mass Index over the Course of Childhood and Adulthood. *PLoS ONE* 2012; **7**: e31470-e31470.

276. Yanovski SZ, Sebring NG. Recorded food intake of obese women with binge eating disorder before and after weight loss. *Int J Eat Disord* 1994; **15**: 135-50.

277. Raymond N, Peterson R, Bartholome L, Raatz S, Jensen M, Levine J. Comparisons of Energy Intake and Energy Expenditure in Overweight and Obese Women With and Without Binge Eating Disorder. *Obesity* 2011;.

278. Black AE, Prentice AM, Goldberg GR, Jebb SA, Bingham SA, Livingstone MB, et al. Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *J Am Diet Assoc* 1993; **93**: 572-9.

279. Prentice AM, Black AE, Coward WA, Davies HL, Goldberg GR, Murgatroyd PR, et al. High levels of energy expenditure in obese women. *British medical journal (Clinical research ed.1981)* 1986; **292**: 983-7.

280. Westerterp K, Goris AHC. Validity of the assessment of dietary intake: problems of misreporting. *Current opinion in clinical nutrition metabolic care* 2002; **5**: 489-93.

281. Mitchell JE, Crow S, Peterson CB, Wonderlich S, Crosby RD. Feeding laboratory studies in patients with eating disorders: a review. *Int J Eat Disord* 1998; **24**: 115.

282. Walsh BT, Boudreau G. Laboratory studies of binge eating disorder. *Int J Eat Disord* 2003; **34 Suppl**: S30-8.

283. Wilson GT. Assessment of binge eating. In: Fairburn CG, Wilson GT (eds). *Binge eating: Nature, assessment, and treatment*. New York, NY, US: Guilford Press, 1993, pp. 227-249.

284. Galanti K, Gluck M, Geliebter A. Test meal intake in obese binge eaters in relation to impulsivity and compulsivity. *Int J Eat Disord* 2007; **40**: 727.

285. Geliebter A, Hassid G, Hashim SA. Test meal intake in obese binge eaters in relation to mood and gender. *Int J Eat Disord* 2001; **29**: 488-494.

286. Anderson DA, Williamson DA, Johnson WG, Grieve CO. Validity of test meals for determining binge eating. *Eating Behav* 2001; **2**: 105-12.

287. Goldfein JA, Walsh BT, LaChausse JL, Kissileff HR, Devlin MJ. Eating behavior in binge eating disorder. *Int J Eat Disord* 1993; **14**: 427-31.

288. Gosnell BA, Mitchell JE, Lancaster KL, Burgard MA, Wonderlich SA, Crosby RD. Food presentation and energy intake in a feeding laboratory study of subjects with binge eating disorder. *Int J Eat Disord* 2001; **30**: 441-446.

289. Sysko R, Devlin M, Walsh BT, Zimmerli E, Kissileff H. Satiety and test meal intake among women with binge eating disorder. *Int J Eat Disord* 2007; **40**: 554-61.

290. Cooke EA, Guss JL, Kissileff HR, Devlin MJ, Walsh BT. Patterns of food selection during binges in women with binge eating disorder. *Int J Eat Disord* 1997; **22**: 187-193.

291. Guss J, Kissileff H, Devlin M, Zimmerli E, Walsh BT. Binge size increases with body mass index in women with binge-eating disorder. *Obes Res* 2002; **10**: 1021-9.

292. Yanovski SZ, Leet M, Yanovski JA, Flood M, Gold PW, Kissileff HR, et al. Food selection and intake of obese women with binge-eating disorder. *Am J Clin Nutr* 1992; **56**: 975-980.

293. Raymond NC, Bartholome LT, Lee SS, Peterson RE, Raatz SK. A comparison of energy intake and food selection during laboratory binge eating episodes in obese women with and without a binge eating disorder diagnosis. *Int J Eat Disord* 2007; **40**: 67-71.

294. Spitzer RL, Devlin M, Walsh BT, Hasin D, Wing R, Marcus M, et al. Binge eating disorder: A multisite field trial of the diagnostic criteria. *Int J Eat Disord* 1992; **11**: 191-203.

295. de Zwaan M, Mitchell JE, Seim HC, Specker SM, Pyle RL, Raymond NC, et al. Eating related and general psychopathology in obese females with binge eating disorder. *Int J Eat Disord* 1994; **15**: 43-52.

296. Adami GF, Gandolfo P, Campostano A, Cocchi F, Bauer B, Scopinaro N. Obese binge eaters: metabolic characteristics, energy expenditure and dieting. *Psychol Med* 1995; **25**: 195-198.

297. Bandini LG, Schoeller DA, Edwards J, Young VR, Oh SH, Dietz WH. Energy expenditure during carbohydrate overfeeding in obese and nonobese adolescents. *Am J Physiol* 1989; **256**: E357-E367.

298. Diaz EO, Prentice AM, Goldberg GR, Murgatroyd PR, Coward WA. Metabolic response to experimental overfeeding in lean and overweight healthy volunteers. *Am J Clin Nutr* 1992; **56**: 641-655.

299. Klein S, Goran M. Energy metabolism in response to overfeeding in young adult men. *Metabolism, clinical and experimental* 1993; **42**: 1201-1205.

300. Cole TJ, Coward WA. Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. *Am J Physiol* 1992; **263**: E965-73.

301. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I Disorders. Patient Edition (SCID-P, Version 2). New York: New York State Psychiatric Institute, Biometrics Research; 1995;.

302. Spitzer RL, Williams JBW, Gibbon M, First MB. *User's guide for the structured clinical interview for DSM-III-R: SCID,* Washington, DC, US: American Psychiatric Association, 1990.

303. Fairburn CG C, Z. The Eating Disorder Examination. New York: Guilford Press; 1993;.

304. Nutritionist IV. In. San Burno: Hearst Corporation; 1999.

305. Raymond NC, Neumeyer B, Warren CS, Lee SS, Peterson CB. Energy intake patterns in obese women with binge eating disorder. *Obes Res* 2003; **11**: 869-879.

306. Bartholome LT, Raymond NC, Lee SS, Peterson CB, Warren CS. Detailed analysis of binges in obese women with binge eating disorder: Comparisons using multiple methods of data collection. *Int J Eat Disord* 2006; **39**: 685-693.

307. Posner BM, Borman CL, Morgan JL, Borden WS, Ohls JC. The validity of a telephone-administered 24-hour dietary recall methodology. *Am J Clin Nutr* 1982; **36**: 546-553.

308. Goldfein JA, Walsh BT, LaChaussee JL, Kissileff HR, Devlin MJ. Eating behavior in binge eating disorder. *Int J Eat Disord* 1993; **14**: 427-431.

309. Guss JL, Kissileff HR, Devlin MJ, Zimmerli E, Walsh BT. Binge size increases with body mass index in women with binge-eating disorder. *Obes Res* 2002; **10**: 1021-1029.

310. Bartholome L, Peterson R, Raatz S, Raymond N. A comparison of the accuracy of self-reported intake with measured intake of a laboratory overeating episode in overweight and obese women with and without binge eating disorder. *Eur J Nutr* 2012;.

311. Grilo CM, White MA, Masheb RM. DSM-IV psychiatric disorder comorbidity and its correlates in binge eating disorder. *Int J Eat Disord* 2009; **42**: 228-234.

312. Azarbad L, Corsica J, Hall B, Hood M. Psychosocial correlates of binge eating in Hispanic, African American, and Caucasian women presenting for bariatric surgery. *Eating Behav* 2010; **11**: 79-84.

313. Jones-Corneille L, Wadden T, Sarwer D, Faulconbridge L, Fabricatore A, Stack R, et al. Axis I Psychopathology in Bariatric Surgery Candidates with and without Binge Eating Disorder: Results of Structured Clinical Interviews. *Obesity Surg* 2010;.

314. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; **4**: 561-571.

315. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* 1988; **56**: 893-897.

316. Stanley MA, Beck JG. Anxiety disorders. *Clin Psychol Rev* 2000; **20**: 731-754.

317. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev* 1988; **8**: 77-100.

318. Muthén LK, Muthén BO. Mplus User's Guide. Fifth Edition. 1998-2007; **5.0**.

319. MacKinnon D, Lockwood C, Hoffman J, West S, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* 2002; **7**: 83-104.

320. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; **51**: 1173-1182.

321. Judd CM, Kenny DA. Process Analysis. *Evaluation Review* 1981; **5**: 602-619.

322. Preacher K, Hayes A. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, computers* 2004; **36**: 717-731.

323. R Development Core Team. R: A language and environment for statistical computing. 2009; **2.9.2**.

324. Peterson C, Thuras P, Ackard D, Mitchell J, Berg K, Sandager N, et al. Personality dimensions in bulimia nervosa, binge eating disorder, and obesity. *Compr Psychiatry* 2010; **51**: 31-36.

325. Fandio J, Moreira R, Preissler C, Gaya C, Papelbaum M, Coutinho W, et al. Impact of binge eating disorder in the psychopathological profile of obese women. *Compr Psychiatry* 2010; **51**: 110-114.

326. Telch CF, Pratt EM, Niego SH. Obese women with binge eating disorder define the term binge. *Int J Eat Disord* 1998; **24**: 313-317.

327. Smyth J, Wonderlich S, Sliwinski M, Crosby R, Engel S, Mitchell J, et al. Ecological momentary assessment of affect, stress, and binge-purge behaviors: day of week and time of day effects in the natural environment. *Int J Eat Disord* 2009; **42**: 429.

328. Smyth J, Wonderlich S, Crosby R, Miltenberger R, Mitchell J, Rorty M. The use of ecological momentary assessment approaches in eating disorder research. *Int J Eat Disord* 2001; **30**: 83.

329. Stein R, Kenardy J, Wiseman C, Dounchis J, Arnow B, Wilfley D. What's driving the binge in binge eating disorder?: A prospective examination of precursors and consequences. *Int J Eat Disord* 2007; **40**: 195-203.

330. Hilbert A, Tuschen-Caffier B. Maintenance of binge eating through negative mood: a naturalistic comparison of binge eating disorder and bulimia nervosa. *Int J Eat Disord* 2007; **40**: 521-530.

331. Haedt Matt A, Keel P. Revisiting the affect regulation model of binge eating: a meta-analysis of studies using ecological momentary assessment. *Psychol Bull* 2011; **137**: 660-681.

332. Pagoto S, Bodenlos J, Kantor L, Gitkind M, Curtin C, Ma Y. Association of major depression and binge eating disorder with weight loss in a clinical setting. *Obesity* 2007; **15**: 2557-2559.

333. Masheb RM, Grilo CM. Examination of predictors and moderators for self-help treatments of binge-eating disorder. *J Consult Clin Psychol* 2008; **76**: 900-904.

334. Lichtman SW, Pisarska K, Berman ER, Pestone M, Dowling H, Offenbacher E, et al. Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. *N Engl J Med* 1992; **327**: 1893-1898.

335. Scagliusi FB, Ferriolli E, Pfrimer K, Laureano C, Cunha CSF, Gualano B, et al. Characteristics of women who frequently under report their energy intake: a doubly labelled water study. *Eur J Clin Nutr* 2009; **63**: 1192-1199.

336. Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res* 2003; **12**: 3-21.

337. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; **289**: 3095-3105.

338. Creed F, Morgan R, Fiddler M, Marshall S, Guthrie E, House A. Depression and anxiety impair health-related quality of life and are associated with increased costs in general medical inpatients. *Psychosomatics* 2002; **43**: 302-309.

339. Chamorro J, Bernardi S, Potenza M, Grant J, Marsh R, Wang S, et al. Impulsivity in the general population: a national study. *J Psychiatr Res* 2012; **46**: 994-1001.

340. Whiteside SP, Lynam DR. The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences* 2001; **30**: 669-689.

341. Mobbs O, Crépin C, Thiéry C, Golay A, Van der Linden M. Obesity and the four facets of impulsivity. *Patient Educ Couns* 2010; **79**: 372-377.

342. Elfhag K, Morey L. Personality traits and eating behavior in the obese: poor self-control in emotional and external eating but personality assets in restrained eating. *Eating Behav* 2008; **9**: 285-293.

343. Nederkoorn C, Smulders FTY, Havermans R, Roefs A, Jansen A. Impulsivity in obese women. *Appetite* 2006; **47**: 253-256.

344. Guerrieri R, Nederkoorn C, Stankiewicz K, Alberts H, Geschwind N, Martijn C, et al. The influence of trait and induced state impulsivity on food intake in normal-weight healthy women. *Appetite* 2007; **49**: 66-73.

345. Davis C, Levitan R, Muglia P, Bewell C, Kennedy J. Decision-making deficits and overeating: a risk model for obesity. *Obes Res* 2004; **12**: 929-935.

346. Davis C, Strachan S, Berkson M. Sensitivity to reward: implications for overeating and overweight. *Appetite* 2004; **42**: 131-138.

347. Davis C, Patte K, Levitan R, Reid C, Tweed S, Curtis C. From motivation to behaviour: a model of reward sensitivity, overeating, and food preferences in the risk profile for obesity. *Appetite* 2007; **48**: 12-19.

348. Guerrieri R, Nederkoorn C, Jansen A. The interaction between impulsivity and a varied food environment: its influence on food intake and overweight. *Int J Obes* 2008; **32**: 708-714.

349. Yeomans M, Leitch M, Mobini S. Impulsivity is associated with the disinhibition but not restraint factor from the Three Factor Eating Questionnaire. *Appetite* 2008; **50**: 469.

350. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *J Psychosom Res* 1985; **29**: 71-83.

351. Bryant, King, Blundell,. Disinhibition: its effects on appetite and weight regulation. *Obesity reviews* 2008; **9**: 409.

352. French S, Epstein L, Jeffery R, Blundell J, Wardle J. Eating behavior dimensions. Associations with energy intake and body weight. A review. *Appetite* 2012; **59**: 541-549.

353. Peluso MAM, Hatch JP, Glahn DC, Monkul ES, Sanches M, Najt P, et al. Trait impulsivity in patients with mood disorders. *J Affect Disord* 2007; **100**: 227-231.

354. Granö N, Keltikangas-Järvinen L, Kouvonen A, Virtanen M, Elovainio M, Vahtera J, et al. Impulsivity as a predictor of newly diagnosed depression. *Scand J Psychol* 2007; **48**: 173-179.

355. Swann A, Steinberg J, Lijffijt M, Moeller,. Impulsivity: differential relationship to depression and mania in bipolar disorder. *J Affect Disord* 2008; **106**: 241.

356. Smith A, Witte T, Teale N, King S, Bender T, Joiner T. Revisiting impulsivity in suicide: implications for civil liability of third parties. *Behavioral sciences the law* 2008; **26**: 779-797.

357. Beskow J. Depression and suicide. *Pharmacopsychiatry* 1990; **23 Suppl 1**: 3-8.

358. Corruble E, Damy C, Guelfi JD. Impulsivity: a relevant dimension in depression regarding suicide attempts? *J Affect Disord* 1999; **53**: 211-215.

359. Kendler KS, Walters EE, Truett KR, Heath AC, Neale MC, Martin NG, et al. Sources of individual differences in depressive symptoms: analysis of two samples of twins and their families. *Am J Psychiatry* 1994; **151**: 1605-1614.

360. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**: 1552-1562.

361. Bezdjian S, Baker L, Tuvblad C. Genetic and environmental influences on impulsivity: a meta-analysis of twin, family and adoption studies. *Clin Psychol Rev* 2011; **31**: 1209-1223.

362. Neale B, Mazzeo S, Bulik C. A twin study of dietary restraint, disinhibition and hunger: an examination of the eating inventory (three factor eating questionnaire). *Twin research* 2003; **6**: 471-478.

363. Choy WC, Lopez-Leon S, Aulchenko YS, Mackenbach JP, Oostra BA, van Duijn CM, et al. Role of shared genetic and environmental factors in symptoms of depression and body composition. *Psychiatr Genet* 2009; **19**: 32-38.

364. Afari N, Noonan C, Goldberg J, Roy Byrne P, Schur E, Golnari G, et al. Depression and obesity: do shared genes explain the relationship? *Depress Anxiety* 2010; **27**: 799-806.

365. Neale, Kendler,. Models of comorbidity for multifactorial disorders. *Am J Hum Genet* 1995; **57**: 935.

366. Mokdad A, Marks J, Stroup D, Gerberding J. Actual causes of death in the United States, 2000. *JAMA (Chicago, Ill.)* 2004; **291**: 1238.

367. Vineis P, Alavanja M, Buffler P, Fontham E, Franceschi S, Gao YT, et al. Tobacco and cancer: recent epidemiological evidence. *J Natl Cancer Inst* 2004; **96**: 99.

368. Klesges RC, Stein RJ, Hultquist CM, Eck LH. Relationships among smoking status, body composition, energy intake, and physical activity in adult males: a longitudinal analysis. *J Subst Abuse* 1992; **4**: 47-56.

369. Perkins KA. Effects of tobacco smoking on caloric intake. *Br J Addict* 1992; **87**: 193-205.

370. Martínez de Morentin PB, Whittle A, Fernø J, Nogueiras R, Diéguez C, Vidal Puig A, et al. Nicotine induces negative energy balance through hypothalamic AMP-activated protein kinase. *Diabetes* 2012; **61**: 807-817.

371. Serdula MK, Byers T, Mokdad AH, Simoes E, Mendlein JM, Coates RJ. The association between fruit and vegetable intake and chronic disease risk factors. *Epidemiology* 1996; **7**: 161-165.

372. Thompson RL, Margetts BM, Wood DA, Jackson AA. Cigarette smoking and food and nutrient intakes in relation to coronary heart disease. *Nutrition research reviews* 1992; **5**: 131-152.

373. Kvaavik E, Meyer H, Tverdal A. Food habits, physical activity and body mass index in relation to smoking status in 40-42 year old Norwegian women and men. *Prev Med* 2004; **38**: 1-5.

374. Chiolero A, Wietlisbach V, Ruffieux C, Paccaud F, Cornuz J. Clustering of risk behaviors with cigarette consumption: A population-based survey. *Prev Med* 2006; **42**: 348-353.

375. Brand J, Chan M, Dowsett M, Folkerd E, Wareham N, Luben R, et al. Cigarette smoking and endogenous sex hormones in postmenopausal women. *J Clin Endocrinol Metab* 2011; **96**: 3184-3192.

376. Field AE, Colditz GA, Willett WC, Longcope C, McKinlay JB. The relation of smoking, age, relative weight, and dietary intake to serum adrenal steroids, sex hormones, and sex hormone-binding globulin in middle-aged men. *J Clin Endocrinol Metab* 1994; **79**: 1310-1316.

377. Steptoe A, Ussher M. Smoking, cortisol and nicotine. *International journal of psychophysiology* 2006; **59**: 228-235.

378. Badrick E, Kirschbaum C, Kumari M. The relationship between smoking status and cortisol secretion. *J Clin Endocrinol Metab* 2007; **92**: 819-824.

379. Koopmans, Slutske, Heath, Neale, Boomsma,. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav Genet* 1999; **29**: 383.

380. Kendler, Neale, Sullivan, Corey, Gardner, Prescott,. A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol Med* 1999; **29**: 299.

381. Batra V, Patkar A, Berrettini W, Weinstein S, Leone F. The genetic determinants of smoking. *Chest* 2003; **123**: 1730.

382. Vink J, Willemsen G, Boomsma D. Heritability of smoking initiation and nicotine dependence. *Behav Genet* 2005; **35**: 397.

383. Maes H, Sullivan P, Bulik C, Neale M, Prescott C, Eaves L, et al. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med* 2004; **34**: 1251.

384. Liu J, Tozzi F, Waterworth D, Pillai S, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010; **42**: 436.

385. AnonymousGenome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010; **42**: 441-447.

386. Thorgeirsson T, Gudbjartsson D, Surakka I, Vink J, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010; **42**: 448.

387. Patterson N, Hattangadi N, Lane B, Lohmueller K, Hafler D, Oksenberg J, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; **74**: 979.

388. Freedman M, Reich D, Penney K, McDonald G, Mignault A, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004; **36**: 388.

Vita

Roseann Elizabeth Peterson was born on June 10th, 1981 in Saint Paul, Minnesota. Following her high school graduation from Roseville Area High School, she attended the University of Minnesota (U of MN) in Minneapolis, Minnesota, where she earned a Bachelor of Arts in Biology, Psychology and a Minor in Chemistry. As part of her Liberal Arts education, she studied the German language and after attending a summer session at Univeristät Heidelberg, she completed her proficiency exam in 2002. During her undergraduate education, she was employed as a research assistant in Dr. Kathleen Thomas's Cognitive Developmental Neuroimaging Laboratory where she administered serial reaction time and implicit learning tasks to children and adults and assisted with electroencephalogram and magnetic resonance imaging. She also taught community education classes such as Crash N' Burn Chemistry and Beginning German at Southwest Public High School. Post-baccalaureate, Ms. Peterson was employed as a research coordinator under the direction of Dr. Nancy C. Raymond, M.D., in the Department of Psychiatry and the Deborah E. Powell Center for Women's Health, A Nationally Designated Center of Excellence. She coordinated studies researching energy intake and expenditure patterns and mechanisms of impulsivity in eating disordered women and organized women's health research conferences. In addition, Roseann volunteered as a crisis counselor at the Rape and Sexual Abuse Center of Minneapolis, MN where she provided crisis intervention and supportive counseling to survivors of sexual violence. She moved to Richmond, Virginia in 2007 to pursue an advanced degree in Human and Molecular Genetics at Virginia Commonwealth University (VCU). Her dissertation advisor was Dr. Hermine H. Maes, Ph.D., Assistant Professor in the Department of Human and Molecular Genetics and Massey Cancer Center. Ms. Peterson studied the genetic and environmental influences on obesity, mood disorders and substance use with an emphasis on statistical methodology. She has received recognition for her work, including a Meritorious Student award from the Society of Behavioral Medicine, the Thompson award for best presentation by a trainee from the Behavior Genetics Association and the Kenneth S. Kendler award for Excellence in Predoctoral Research. She has served the VCU graduate student community as Research Symposium Chair and Vice President of the Graduate Student Association. She will continue her research in Richmond, Virginia as a postdoctoral fellow at the Virginia Institute for Psychiatric and Behavioral Genetics at VCU.