Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Helene Fevrier Date

Polygenic Risk Scores and Schizophrenia: Association with Gender and Age of Onset in an Ashkenazi Jewish Population

By

Helene Fevrier
MSPH

Department of Epidemiology

Jennifer G. Mulle, MHS, PhD

Committee Chair

Polygenic Risk Scores and Schizophrenia: Association with Gender and Age of Onset in an Ashkenazi Jewish Population

By

Helene Fevrier

Bachelor of Arts

Boston University

2014

Thesis Committee Chair: Jennifer G. Mulle, MHS, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology
2016

Abstract

Polygenic Risk Scores and Schizophrenia: Association with Gender and Age of Onset in an Ashkenazi Jewish Population

By Helene Fevrier

Schizophrenia (SZ) is a severe and debilitating psychiatric disorder with a range of symptoms and outcomes that range from remission to chronic states where treatment options are not effective. Remission in SZ is association with female gender, later age of onset and good social functioning prior to onset. Twin studies have shown 80% heritability between twins and 60% heritability in family studies. SZ is a polygenic disorder, where the individual contribution of each SNP is minimal, and the pooled effect of many mutations contributes to a higher risk of SZ and other psychiatric disorders, including bipolar disorder (BP). The data for this study was from the Epidemiology Genetics Program at the Johns Hopkins University School of Medicine, where SZ cases and unaffected controls of Ashkenazi Jewish descent were ascertained. Genome-wide SNP data for 600 SZ cases, 446 BP cases, and 508 controls was used to create risk scores for each individual at 3 different association thresholds. Risk scores were investigated for associations between gender and age of onset. Risk score was associated with SZ phenotype after adjusting for sex (OR: 1.20, 1.03-1.40). In females, higher scores may be related to earlier ages of onset; no such relationship was found in men. Risk scores for BP cases were created using a dataset of SZ cases and controls to test whether polygenic risk score, but not gender, was associated with BP phenotype ($p<0.0001$). Age of onset in BP cases was not associated with risk score. These findings support previous literature that gender is associated with genetic risk for SZ, and that SZ and BP may share common risk alleles.

Polygenic Risk Scores and Schizophrenia: Association with Gender and Age of Onset in an Ashkenazi Jewish Population

By

Helene Fevrier

Bachelor of Arts

Boston University

2014

Thesis Committee Chair: Jennifer G. Mulle, MHS, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology
2016

Acknowledgements

I would like to offer my great appreciation to Dr. Jennifer Mulle for her continued support and mentorship.

I would also like to thank the Epidemiology Genetics Program at the Johns Hopkins University School of Medicine.

Table of Contents

Introduction

Schizophrenia (SZ) is a complex and debilitating disorder that affects 1% of the population worldwide (Chen, Cao et al. 2015). This psychiatric disorder is characterized by positive symptoms (including delusions, hallucinations, and disorganized behavior or speech) and negative symptoms (including alogia, avolition, and asociality) (Association and Association 2000). Alogia is a lack of speech while avolition is characterized by decreased motivation to perform routine activities such as attending school, continuing employment and engaging in social activities. These are classified as negative symptoms because there is a deficiency in normal functioning. Most individuals with SZ display a combination of both positive and negative symptoms (Elis, Caponigro et al. 2013, Lang, Kösters et al. 2013). A diagnosis is made after a combination of these symptoms has been present for greater than six months. Individuals who are diagnosed usually have a difficult time with self-care, interpersonal relationships, and maintaining an occupation.

SZ is a heterogeneous disorder (Chen, Cao et al. 2015). Longitudinal studies have reported a range of symptoms and severity. Outcomes can vary from remission after the first psychotic episode to severe and chronic states where treatment options are not effective (Lang, Kösters et al. 2013). Remission in non-psychiatric diseases is generally defined by a lack of symptoms and a return to baseline health. However, in psychiatric disorders, remission is not the absence of all symptoms; rather, it is defined by diminished severity of symptoms with only mild disability. (Andreasen, Carpenter Jr et al. 2014). A patient with SZ may be in remission when symptoms no longer significantly interfere with day to day functioning. Several factors are associated with remission

including female gender, a later age of onset, and good social functioning prior to onset (Haro, Novick et al. 2006). SZ patients may also experience relapses after achieving remission; Haro et al. estimated that 25% of patients who achieved remission later relapsed. A relapse is defined by a worsening of symptoms, an increased number of hospitalizations or a decrease in social functioning (Haro, Novick et al. 2006). Higher risk of relapse is associated with substance abuse and hostile behavior. In a study of males with schizophrenia, patients who had not achieved remission one year after onset, had also not achieved remission after 4 and 7 years (Ceskova, Prikryl et al. 2011). Negative symptoms, male gender and longer periods of untreated psychosis are associated with poorer outcomes in patients.

There are considerable differences in symptoms and outcomes between males and females with SZ. Males have an earlier age of onset, while onset in females is general 3-5 years after males (Leung and Psych 2000, Lindamer, Lohr et al. 2014). However, no significant difference have been observed in mean age of onset in childhood-onset SZ between males and females (Ochoa, Usall et al. 2012). Males also have a higher incidence of SZ; the odds ratio for SZ for males compared to females in 1.4 (McGrath, Saha et al. 2004). The range of symptoms may also vary between sexes. Although studies have shown mixed results, it had been suggested that men have more negative or disorganized symptoms, while women with SZ have symptoms of depression and anxiety (Ochoa, Usall et al. 2012, Goldstein, Cherkerzian et al. 2013). Outcomes between males and females also exhibit significant differences. The rate of relapse is higher in males, while the rate of remission is higher in females (Ochoa, Usall et al. 2012).

Environmental risk factors are believed to play a role in the risk of SZ. Environmental risk factors can increase risk before birth, during early childhood and during adolescent and early adult years. The relationship between environmental risk factors and stage of life may contribute to increased risk or range of symptoms that are seen in SZ (Cannon, Jones et al. 2002). It is hypothesized that paternal age over 40 years is associated with SZ through a mechanism of passing a greater number of de novo mutations from father to offspring. (Matheson, Shepherd et al. 2011, Harper, Towers-Evans et al. 2015). As age increases, the number of de novo mutations increases, however, no association has been determined for increased maternal age and risk of SZ (Kong, Frigge et al. 2012).

Low birth weight (<2000g) has also been associated with increased risk of SZ in many, but not all studies of low birth weight. Low birth weight may be a risk factor for many psychiatric diseases (BP and substance abuse), and is not limited to SZ (Harper, Towers-Evans et al. 2015). Maternal diabetes and season of birth have also shown small effect sizes of increased risk for SZ (Matheson, Shepherd et al. 2011). Risk factors that may contribute to SZ risk and symptoms later in life include cannabis use and substance abuse, and urbanicity (Matheson, Shepherd et al. 2011). There are a wide variety of environmental risk factors that are associated with SZ. Not only do these risk factors vary by individual, but the interactions between environmental and genetic risk factors are complex but very important in determining risks for SZ.

Although associations between certain environmental risk factors and SZ have been determined, much of the risk of SZ is heritable. SZ is a genetic disorder with high familial heritability. Twin studies have shown 80% heritability between twins, and 60% heritability in family studies (Sullivan, Kendler et al. 2003). It is estimated that having an affected 1st degree relative increases the risk of SZ 10-fold (Gottesman and Shields 1982). However, there is no evidence that inheritance of SZ follows a Mendelian pattern (Owen, Craddock et al. 2010, Sullivan, Daly et al. 2012).

Recently, significant advances have been made in discovering genetic factors associated with an increased risk of SZ. These include copy number variants (CNV), where large portions, (greater than 100,000 bases) of the genome have been deleted or duplicated, and common variants, which are smaller in size and occur more frequently in the genome (Stone, O'Donovan et al. 2008). CNVs are rare and occur in less than 1% of the population (NUMBER and CASES 2010). The number of genes that CNVs span is a better predictor of increased SZ risk than simply the number of CNVs in an individual's genome. Additionally, many of the CNVs that have been associated with increased risk of SZ have also been associated with increased risk of other psychiatric disorders including, autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) (Sullivan, Daly et al. 2012).

Genome wide association studies (GWAS) have also prompted new discoveries about the genetics of SZ. Recent advances in GWAS have identified a number of single nucleotide polymorphisms (SNPs) which are associated with an increased risk of SZ. SZ is a highly

polygenic disorder; although the individual contribution of each SNP is very small, when the effect of these mutations is pooled, they contribute to a higher risk of SZ. It is hypothesized that the heterogeneity of the disease and its outcomes comes from the combination of risk from many small associated SNPs and rarer, large CNVs. A number of these SNPs show not only an association for increased risk of SZ but also are also associated with an increased risk of bipolar disorder (BP) (Sullivan, Daly et al. 2012). Gottesman et al. studied the offspring of SZ and BP affected parents. The 2010 study found the risk of SZ to be 27.3% higher when offspring has two affected parents, 7.0% where offspring had one affected parents and 0.86% in couples where neither parents was affected (Gottesman, Laursen et al. 2010). This trend is also seen among offspring of parents affected by BP, where the risk of BP in offspring with two affected parents, one affected parent and no affected parent is 24.9%, 4.4% and 0.48% respectively. In addition, offspring of two parents with BP had a cumulative incidence of SZ approximately four times higher than the general population (Gottesman, Laursen et al. 2010). This study supports the hypothesis that SZ is highly heritable, and that SZ may share common risk alleles with BP.

Despite the large effect size of CNVs, it is hypothesized that much of the variance in heritability comes from smaller common variants. Through recent GWAS many associated risk alleles have been identified. The variants that are detected by GWAS are common, meaning that the allele frequency is found in greater than 1% of the population (Chen, Cao et al. 2015). However, large samples sizes are required to detect significant alleles; smaller sample sizes do not have adequate power to do so (Owen, Craddock et al.

2010). One of the early GWAS's used 479 cases and 2,937 controls and investigated 12 loci with a moderate association, and 3 with strong associations. The strongest association for a risk allele was found in the Zinc Finger Protein 804A (ZFP804A) region (O'donovan, Craddock et al. 2008). Ripke et al. found 22 significantly associated loci, of which 13 were novel findings, using a sample size of 5,001 cases and 6,243 controls (Ripke, O'Dushlaine et al. 2013). This study provided evidence for the role of the major histocompatibility complex (MHC) in development of SZ (Schwab and Wildenauer 2013). The latest findings were published by the Schizophrenia Working Group of the Psychiatric Genomics Consortium. Using a sample size of 150,064 individuals, 108 loci were found to be associated with SZ (Consortium 2014). Of the 108 associated loci, 75% were found in protein coding regions, and an additional 8% were found within 20KB of a protein coding region (Consortium 2014). GWAS studies are needed to study the small effect sizes of individuals SNPs because the additive effects of these alleles confer an increased risk of SZ.

GWAS has been particularly successful at identifying variants related to psychiatric disorders because they are capable of screening thousands of individuals. GWAS can be a helpful tool in detecting risk alleles in psychiatric disorders because evidence has shown many of the same risk alleles may confer risk to multiple psychiatric disorders including, BP, ASD, ADHD, major depressive disorder (MDD) and SZ. In 2009, Lichtenstein et al. reported findings that when parents were diagnosed with BP, the risk of SZ increased for offspring of affected parents (RR=2.4 95%CI: 2.1-2.4). Furthermore, when a sibling was diagnosed with BP, other siblings were at higher risk for being diagnosed with SZ

(RR=3.9, 95%CI: 3.4-4.4) (Lichtenstein, Yip et al. 2009). Similarly, offspring of parents diagnosed with SZ had an increased relative risk for BP (RR=5.2 95%CI: 4.4-6.2); while siblings diagnosed with SZ, increased with relative risk of BP diagnosis among other siblings (RR=3.7, 95%CI: 3.2-4.2) (Lichtenstein, Yip et al. 2009). Additionally, the Cross-Disorder Group of the Psychiatric Genomics Consortium studied 33,332 cases and 27,888 controls with one of five psychiatric disorders (ASD, BP, SZ, ADHD, and MDD) (Consortium 2013). The study identified 4 loci, two of which implicated calcium-voltage gated channels, which were associated with an increased risk of these 5 disorders. This was the first genome-wide study that suggested a shared genetic risk between these five disorders (Consortium 2013). This novel finding might explain the simultaneous occurrence of these disorders and suggests that voltage gated calcium channels are an important pathway in the development of disease.

Significant advances in psychiatric genetics have been made by recent GWAS discoveries. However, GWAS requires very large sample sizes to detect associations between risk alleles and potential outcomes. Additionally, there are many alleles that do not reach significance in these large studies. In polygenic disorders, specifically SZ and other psychiatric disorders, increased risk is attributed to large mutations (CNVs) and the collective contribution of smaller SNPs that may not reach significance on their own. Polygenic risk scores summarize the combined effect of these alleles and their association to the outcome. A sum of each allele that is protective against or increases

risk of the outcome can calculated and applied as a risk score for each individual in a dataset (Musliner, Seifuddin et al. 2015). This can be used to find genetic associations well as predict risks.

Polygenic risk scores can be used to measure the collective effect of genetic markers that may be associated but do not reach genome-wide statistical significance in GWAS studies (Dudbridge 2013). Although these genetic markers alone may not have been detected, their combined effect can be associated with a disease outcome. A polygenic risk score incorporates many genetic markers into a single score by ranking GWAS results in order of significance using certain p-value thresholds. Subsequently, an analysis is run on a discovery dataset to identify SNPs associated with the outcome of interest (protective or harmful). Once each allele is given a measure of association, a polygenic risk score can be calculated in a target dataset, where a weighted sum of the associated alleles is applied to each individual in the dataset (Dudbridge 2013, Musliner, Seifuddin et al. 2015). Polygenic risk scores can then be applied to find evidence of genetic effects from markers that alone are not significant, and to investigate shared genetic contribution in related disorders. This was done by the International Schizophrenia Consortium in 2009 which found genetic similarities between SZ and BP (Purcell, Wray et al. 2009, Dudbridge 2013). Currently, the most common use for polygenic risk scores involves testing genetic associations; however, as GWAS sample sizes grow, risk scores will have more power to predict risk for outcomes of interest (Dudbridge 2013).

The use of polygenic risk scores has been exceptionally useful in studying psychiatric disorders. In particular, Purcell et al. performed a GWAS of 3,322 SZ cases and 3,857

controls, and used risk scores to provide evidence that SZ is polygenic, and thousands of non-significant alleles could collectively increase risk of SZ. In the analysis, an independent sample was used to determine the relationship between SNPs and SZ. Using a subset of 74,062 SNPs, individuals in the target dataset were given a score. In this sample, higher polygenic risk scores were associated with SZ diagnosis; a higher mean polygenic risk score was seen in cases when compared to controls. In addition, in two independent samples, a polygenic risk score for SZ was associated BP diagnosis. This overlap suggests that BP and SZ might have shared genetic components. To verify this finding, Purcell et al. looked at the mean scores in cases and controls of non-psychiatric illnesses including coronary artery disease, type I and type II diabetes, hypertension, Crohn's disease and rheumatoid arthritis. When mean scores were compared to non-psychiatric diseases, there was no association between score and disease diagnosis (Purcell, Wray et al. 2009)

Ahn et al. also demonstrated that higher polygenic risk score comparisons were associated with cases diagnosed with childhood onset SZ, when compared to full healthy siblings. Using the 80 most significantly related variables from the 108 risk loci identified by the Schizophrenia Working Group of the Psychiatric Genomics, 130 cases who met DSM-IV criteria for SZ before the age of 13, were compared 103 to full healthy siblings (Ahn, An et al. 2014, Consortium 2014). When the polygenic risk scores for cases and controls were compared, individuals with childhood onset SZ had significantly higher scores than their siblings ($p< 4x10^{-5}$). In addition, Ahn et al. provided further evidence that some psychiatric disorders may share similar genetic components. Using a set of

SNPs published by the Psychiatric Genetic Consortium for ASD, an association between score and SZ cases was significant at the GWAS level. This finding supports the theory that this related set of disorders may share common markers.

Although many SNPs do not reach significance in GWAS and other large studies, there is sufficient evidence that the collective effect of these common variants may play an important role in risk of SZ (Purcell, Wray et al. 2009, Ahn, An et al. 2014). These common variants in combination with larger CNVs explain a large portion of the genetic contribution to SZ risk. In the proposed analysis, two association analyses will be conducted with a population of 600 SZ cases and 508 controls. One will consist of a discovery and target population created from the 1108 cases and controls, while the other will use all SZ cases and controls as both the discovery and target dataset. Risk scores will be created from the association analysis, for each individual, and the association between polygenic risk score and case control status by gender and age of onset will be studied. In addition, a discovery dataset of SZ cases and controls will be used to create risk scores in a target dataset of BP cases and controls to study case control status by gender and age of onset.

Methods

Data Source

Cases and controls were ascertained through the Epidemiology Genetics Program at the Johns Hopkins University School of Medicine. Cases were identified between 1996-2007 using national advertisements in Jewish periodicals, the study website and Jewish community and mental health organizations. Cases were interviewed by PhD clinicians

using the Diagnostic Interview for Genetic Studies and diagnoses were made by 2 research clinicians based on DSM-IV diagnostic criteria (Nurnberger, Blehar et al. 1994). Family history of affected individuals was recorded. Controls were selected from 2003-2007 and were screened for any lifetime psychotic illnesses. Further details on recruitment of cases and controls have been previously published (Fallin, Lasseter et al. 2005, Goes, McGrath et al. 2015). The final sample contained 1554 observations: 600 SZ cases, 446 BP cases and 508 controls.

Genetic Data

Plink (Plink 1.07) was used to investigate SNPs and calculate summary statistics. The original sample contained 364,227 markers and had a genotyping rate of 0.99826 (Purcell, Neale et al. 2007). No Mendelian errors were detected at error rates above 0.1. Two markers were excluded based on Hardy Weinberg Equilibrium at $p<0.00001$. No individuals were removed for low genotyping. 706 SNPs were removed for failing the allele frequency test with minor allele frequency greater than 0.01. No individuals were removed for low genotyping. The final sample contained 363,519 SNPs and 1554 individuals; 600 SZ cases, 446 BP cases, and 508 controls.

Association and Risk Scoring of Target Dataset

Discovery and target datasets were created by randomly selecting half of all observations for each dataset. Frequency distributions of sex, phenotype, affection, and age of onset were verified by chi squared or t-tests, and were similar in both the discovery and target datasets. To ensure that randomization was successful in the discovery and target

datasets, a second randomization for a new discovery and target dataset was created. BP cases were removed from the discovery and target datasets to analyze only SZ cases and controls. An association analysis of the discovery dataset was performed in Plink to compare allele frequencies between SZ cases and controls. Three p-value thresholds were selected for generating polygenic risk scores: (thresholds). The natural log of the odds ratio (OR) from the association analysis for any SNP that had an associated p-value of <0.0001, <0.001 or <0.01, was used to create a polygenic risk score in the target dataset. Each p-value threshold included all SNPs significant at the p-value; the risk scores created from SNPs with a p-value <0.01, also included all SNPs that were significant at <0.0001 and <0.001. One risk score was output for each individual in the target dataset; this was repeated for each p-value threshold.

Risk scores from the target dataset were categorized into quartiles. Risk of case status by quartile was analyzed using the lowest quartile as the reference group. All analyses were performed at each p-value threshold. Univariate logistic regression in SAS (SAS Institute, Cary NC) was used to analyze the association between case-control status and sex or score. Multivariate logistic regression was performed to analyze the association between case-control status and score, sex, and any possible interactions. Additionally, among the cases, risk score was analyzed as a predictor of age of onset for the entire population, stratified by sex.

Secondary Analysis

In addition to creating risk scores for the target dataset, an association analysis was performed on all SZ cases and controls as a means to increase power in the study. The natural log odds of any SNP that had a p-value of <0.0001, <0.001 or <0.01 from the association analysis was used to create risk scores for all SZ cases and controls. One score was output for each individual in the dataset at each p-value threshold. The same analyses from the target dataset were applied to the entire dataset.

Bipolar Analysis

In all previous analyses BP cases were removed to study associations between SZ cases and score, sex, and age of onset. Previous studies have shown that SZ and BP may share common risk alleles. We investigated this by creating a discovery dataset with all SZ cases and controls and a target dataset with only BP cases and controls. The natural log odds for each SNP with a p-value of <0.0001, <0.001, or <0.01 was used to create risk scores in the target dataset of BP cases and controls. Risk scores were categorized into quartiles. Quartiles were compared, using the lowest quartile as the reference group, to determine increases in risk of case status by score. Univariate and multivariate logistic regression were used to determine if BP cases were associated with higher scores or sex. Interaction between sex, score quartiles and score as a continuous variable were evaluated. Age of onset for BP cases was also investigated to determine any association between risk score and age of onset. All analyses were conducted in SAS 9.4 (SAS Institute, Cary NC).

Results

Data Characteristics

The majority of the sample was male (53.7%, n=834) (Table 1). Of individuals affected by SZ, 66.8% were male, while 43.9% of controls were male. The mean age of onset for SZ was 20.1 for males and 20.4 for females; the mean age of onset for BP was 22.7 for males and 24.7 for females. The sample contained genotypes for 364,227 SNPs.

SZ Analysis

The discovery and target datasets created each contained 777 observations. Once BP cases were removed 554 observations remained. There were no significant differences between frequencies of sex, age of onset and affection type at alpha = 0.05 (Table 2). A second discovery and target dataset was created to confirm that randomization was successful. Chi-squared tests, and t-tests confirmed that there were no significant differences between the second discovery and target datasets. In the discovery dataset 36 SNP's were significant at the p-value threshold <0.0001, 343 SNPs were significant at p<0.001 and 3842 were significant at p<0.01 (Table 3). Analyses were performed for the target dataset for each p-value threshold. The mean score of the SZ SNPs at p<0.0001 was -0.033, -0.008 at <0.001 and -0.0006 at p<0.01, while the mean score of the controls were -0.037 p<0.0001, -0.0105 and <0.001 and -0.00149 at p<0.01. The only statistically significant difference between scores of the affected and unaffected were at p<0.01 using alpha = 0.05.

Score quartiles were compared using the lowest quartile as the reference group. In the target dataset there was a trend of increasing odds per score quartile. At the p-value threshold $p<0.01$, when comparing the highest quartile to the lowest quartile, the odds of PZ phenotype increase by 86% (Table 4). Score as a continuous variable, and as a categorical variable in quartiles, was associated with phenotype in a univariate logistic regression model at $p<0.01$ (odds ratio 1.20 [95% confidence interval 1.03, 1.39]) (Table 5). Quartiles of score were associated with higher odds of SZ at both <0.0001 (OR: 1.07, 0.92-1.25) and $p<0.001$ (OR: 1.11, 0.96-1.29), but were not statistically significant. Female sex was protective for SZ case status in a univariate model (OR: 0.42, 0.29-0.59). Interaction terms between score and sex were assessed in a multivariate model but were removed from the model because they were not significant. The final multivariate model included score as a categorical variable and adjusted for sex. This model was repeated at all three p-value thresholds. At $p<0.0001$ and $P<0.001$, higher quartiles and male sex was associated with an SZ phenotype, however these estimates were not statistically significant (OR: 1.06, 0.91-1.23; OR: 1.13, 0.97-1.31, respectively). At $P<0.01$, higher quartiles of score and male sex were significantly associated with SZ phenotype (OR: 1.20 [1.03, 1.40]) (Table 6). In the analysis of age and score among individuals with SZ phenotype, there was a trend where higher scores were associated with earlier ages of onset, however this trend was only seen in females $p<0.0001$ and $p<0.001$, and was not significant (Table 7).

Creating independent target and discovery datasets for the association analyses and score creation limited the power of the analyses. To investigate associations of score, sex and

age within the entire dataset, an association analysis was performed using all SZ cases and controls. A statistically significant difference of mean scores was observed for all p-value thresholds (Table 3). A comparison of the highest and lowest quartiles determined a significant association between score quartile and SZ phenotype (Table 8). In a univariate logistic analysis, score was significantly associated with phenotype at $p<0.0001$ and $p<0.001$ (Table 5). Quartiles of score followed this same pattern. At $p<0.0001$, the OR for score quartiles was 2.25 (2.22, 2.87), and the OR at $p<0.001$, was 34.45 (22.53, 52.80). The final multivariate model for all SZ cases and controls included score quartiles and sex. At $p<0.0001$ the adjusted OR for phenotype was 2.47 (2.17-2.82) (Table 6). In the univariate analysis of age of onset among SZ cases, there was no difference in age of onset observed between males and females. However, higher scores were associated with earlier onset ages among females at $p<0.0001$ and $p<0.001$ (Table 7).

Bipolar Analysis

Association analyses were run on a discovery dataset of SZ cases and controls and applied to a target dataset of BP cases and controls to create risk scores. Risk scores for BP cases, at each p-value threshold, were significantly different from controls risk scores. Using the first quartile as the reference group, each subsequent quartile was associated with a higher risk of BP phenotype. When comparing the 4th quartile, to the first quartile, the OR associated with BP phenotype is 4.06 (2.77, 5.95) (Table 9). This is also observed in a univariate logistic regression model with phenotype and score. Continuous score, as well as score quartiles, is associated with increased odds of BP phenotype. At $p<0.0001$

there is a 58% increase in odds, per increase in score quartile (OR: 1.56, 1.40-1.78) (Table 4). This significant increase in odds is also reported at $p<0.001$ (OR: 4.19, 3.52-4.99) and $p<0.01$ (OR: 71.53, 40.02-127.86). The univariate model of phenotype and score quartiles was the final model because neither age, sex, nor any interactions were significantly associated with BP phenotype. Age of onset for BP cases did not differ by score, however, age of onset did differ by sex, where females had a higher age of onset than males (24.7 years and 22.7 years, respectively) (Table 6).

Discussion

SZ is a complex and heterogeneous disorder with a wide range of symptoms and severity. As expected, the majority of SZ cases were male (66.8%), and the OR associated with male gender in our sample was 2.57 (2.02, 3.29). This association estimate, for male gender and SZ phenotype, is slightly higher than previously reported OR of 1.4 (McGrath, Saha et al. 2004), but consistent with findings that males are at slightly higher risk for SZ. . The distribution of males and females in controls and BP cases were similar (43.9% and 47.1%, respectively). We found no significant difference in mean age of onset for SZ cases, despite the typical age of onset for SZ in women being approximately 5 years later, on average, then in men. A difference in mean age of onset between males and females was seen for BP cases (22.7 years for males vs. 24.7 for females). This is likely a characteristic of our sample rather than a trend for age of onsets of SZ and BP.

In the target dataset, the mean score differences between affected SZ cases and unaffected controls are statistically significant at $p<0.01$ in the univariate analysis. The

OR's for score quartile and SZ phenotype at the $p<0.01$ threshold is 1.20 (1.03, 1.39), while the OR's for the p-value thresholds of $p<0.0001$ and $p<0.001$ are 1.07 and 1.11, however are not significant. The association at $p<0.01$ is significant, while the associations at $p<0.001$ and $p<0.01$ are not significant but show an increasing trend; each increase in score quartile is associated with increased odds of SZ phenotype. This non-significant trend is likely because the sample is underpowered as the total sample was split into 2 datasets for this analysis. This trend is also seen in the entire SZ dataset, where mean differences between affected and unaffected individuals are statistically significant at all p-value thresholds. In addition, in the entire dataset the score quartile OR's increase dramatically as the p-value thresholds get larger in the univariate analysis. The OR for score quartiles in the entire dataset at $p<0.0001$ is 2.52, while the OR at $p<0.001$ is 34.49. This gives evidence that with a larger sample size with increased statistical power, a significant association for the target dataset would be seen at lower p-value thresholds. Additionally, when the highest quartile was compared to the lowest score quartile in a univariate analysis, increased odds of SZ phenotype was observed at $p<0.0001$ and $p<0.001$. Although these were not significant within the target dataset, when all SZ cases and controls were used each score quartile above the reference group, at each p-value threshold, was significantly associated with increased odds of SZ. This indicates that at $p<0.0001$ and $p<0.001$, the target dataset was underpowered. At $p<0.01$, in the target dataset, comparing the highest score quartile to the lowest score quartile gives increased odds of 1.86 (1.15, 3.01).

The target dataset also showed a trend of increased odds of SZ per score quartile in the multivariate analysis. At $p<0.0001$ and $p<0.001$ the association was not significant, however, at $p<0.01$ a one-unit increase in score quartile, adjusting for sex, was associated with a 20% increased odds of SZ phenotype. It is likely that this association was not significant at $p<0.0001$ and $p<0.001$ because the small sample size led to an underpowered association. However, when multivariate regression was run on the entire dataset, the association between score quartile and SZ phenotype, adjusting for sex, was significant at $p<0.0001$ and $p<0.001$. At $p<0.0001$, each increase in score quartile was associated with 2.47 times the odds of SZ phenotype. At $p<0.001$, each increase in score quartile was associated with 33.47 times the odds of SZ phenotype.

There is a known association that SZ onset in women generally occurs at later ages than in men. We investigated this trend to see if high scores were associated with earlier age of onsets. Although no difference in age of onset between men and women existed in a univariate model, we observed a trend that higher scores in women were associated with younger ages of onsets in our target dataset (Figure 1). This trend was not seen in men. In the whole dataset, at the thresholds $p<0.0001$ and $p<0.001$ there was a significant trend among females, that higher scores were associated with younger ages of onset. This supports the trend seen in the target dataset, and suggests that the trend may not have reached significance because it was underpowered (Figure 2). No such trends were seen in males, or between males and females among BP cases.

It has been previously reported that SZ and BP may share common risk alleles. To investigate this in our sample, we created discovery dataset of SZ cases and controls and a target dataset of BP cases and controls. Using the lowest quartile as the reference quartile, each subsequent quartile is associated with an increased odds of BP phenotype, using SZ scores as predictors. The odds of BP phenotype increase 4.06 times when the highest quartile is compared to the lowest quartile at $p<0.0001$. These odds increase to 91.75 when the p-value threshold $p<0.001$ is used. No odds ratio could be calculated for $p<0.01$ because there were no BP cases in the lowest quartile, so it could not be used for comparison. This agrees with previous research that SZ may share common risk alleles with BP. These findings match the univariate logistic regression. Continuous scores for BP cases all showed statistical significance, however, the OR's for continuous score were very high, likely due to the very high correlation between SZ scores and BP phenotype. For this reason, scores were categorized into quartiles. Score quartiles at $p<0.0001$, $p<0.001$ and $p<0.01$ had odds ratios which were significantly associated with BP phenotype. At $p<0.0001$, an increase of one SZ score quartile was associated with 1.58 times the odds of BP phenotype (1.40, 1.78). At $p<0.001$, the odds of BP phenotype increased by 4.19, and at $p<0.01$, the odds of BP phenotype increased 71.53. A multivariate logistic regression was performed to adjust for sex, however sex was found to not be an important predictor of BP phenotype and the final model was left as the univariate model.

Strengths of the study include the use of genome-wide data; the sample contained genotypes for 364,227 SNPs and a genotyping rate of 0.998. All SZ cases, BP cases and

controls were selected from populations of Ashkenazi Jewish descent which allowed for consistency in case evaluation and ensured that SZ cases were distinct from BP cases. Additionally, selecting cases and controls from a single genetic background limits the influence of population stratification.

Another strength of the study was the increasing strength of the analysis. Each univariate and multivariate analysis was performed three times, on a dataset where scores were created from SNPs significant at $p<0.0001$, $p<0.001$ and $p<0.01$. This allowed us to test our hypothesis at varying significances and determine that trends were seen in all dataset and were not associations of chance. Using a separate discovery and target datasets of difference SZ cases and controls in addition to using a discovery and target dataset of all SZ cases and controls allowed us to compare trends between risk score and SZ phenotype despite the small sample size.

To create the discovery and target datasets, the sample size was split in half. Many trends between scores and SZ phenotype, and scores and age of onset were seen at lower p-values, but were not statistically significant because they were underpowered due to the small sample size. To investigate these results, all SZ cases and controls were used as the discovery and target dataset. The univariate and multivariate analyses of the entire dataset followed the same trends seen in the target dataset, but were all statistically significant. This high degree of association was expected because the same cases and controls were used to make scores, however this allowed us to confirm trends seen in the target dataset that may have been underpowered. Additionally, the sample includes cases and controls

from Ashkenazi Jewish descent which may limit the applications of these findings to other populations. Another limitation of this study is that ages of onset were not available for each SZ or BP case. Approximately 17% of cases (SZ or BP) had missing information for age of onset. Even with these limitations this analysis provides insight into the use of risk scores to evaluate risk of SZ, and how SZ risk alleles may also confer a higher risk of BP. There is biological plausibility for why females with higher scores have earlier ages of onset for SZ, however further research is needed to determine is this trend is seen in other populations.

References:

Ahn, K., et al. (2014). "Common polygenic variation and risk for childhood-onset schizophrenia." Molecular psychiatry.

Andreasen, N. C., et al. (2014). "Remission in schizophrenia: proposed criteria and rationale for consensus." American Journal of Psychiatry.

Association, A. P. and A. P. Association (2000). Diagnostic and statistical manual-text revision (DSM-IV-TRim, 2000), American Psychiatric Association.

Cannon, M., et al. (2002). "Obstetric complications and schizophrenia: historical and meta-analytic review." American Journal of Psychiatry.

Ceskova, E., et al. (2011). "Outcome in males with first-episode schizophrenia: 7-year follow-up." The World Journal of Biological Psychiatry **12**(1): 66-72.

Chen, J., et al. (2015). "Genetic studies of schizophrenia: an update." Neuroscience bulletin **31**(1): 87-98.

Consortium, C.-D. G. o. t. P. G. (2013). "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis." The Lancet **381**(9875): 1371-1379.

Consortium, S. W. G. o. t. P. G. (2014). "Biological insights from 108 schizophrenia-associated genetic loci." Nature **511**(7510): 421-427.

Dudbridge, F. (2013). "Power and predictive accuracy of polygenic risk scores." PLoS Genet **9**(3): e1003348.

Elis, O., et al. (2013). "Psychosocial treatments for negative symptoms in schizophrenia: current practices and future directions." Clinical psychology review **33**(8): 914-928.

Fallin, M. D., et al. (2005). "Bipolar I disorder and schizophrenia: a 440–single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios." The American Journal of Human Genetics **77**(6): 918-936.

Goes, F. S., et al. (2015). "Genome-wide association study of schizophrenia in Ashkenazi Jews." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **168**(8): 649-659.

Goldstein, J. M., et al. (2013). "Sex differences in the genetic risk for schizophrenia: History of the evidence for sex-specific and sex-dependent effects." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **162**(7): 698-710.

Gottesman, I. and J. Shields (1982). "Schizophrenia: The Epigenetic Puzzle Cambridge Univ." Press, Cambridge.

Gottesman, I. I., et al. (2010). "Severe mental disorders in offspring with 2 psychiatrically ill parents." Archives of general psychiatry **67**(3): 252-257.

Haro, J. M., et al. (2006). "Remission and relapse in the outpatient care of schizophrenia: three-year results from the Schizophrenia Outpatient Health Outcomes study." Journal of clinical psychopharmacology **26**(6): 571-578.

Harper, S., et al. (2015). "The aetiology of schizophrenia: what have the Swedish Medical Registers taught us?" Social psychiatry and psychiatric epidemiology: 1-9.

Kong, A., et al. (2012). "Rate of de novo mutations and the importance of father/'s age to disease risk." Nature **488**(7412): 471-475.

Lang, F., et al. (2013). "Psychopathological long-term outcome of schizophrenia–a review." Acta Psychiatrica Scandinavica **127**(3): 173-182.

Leung, M. and C. M. Psych (2000). "Sex differences in schizophrenia, a review of the literature." Acta Psychiatrica Scandinavica **101**(401): 3-38.

Lichtenstein, P., et al. (2009). "Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study." The Lancet **373**(9659): 234-239.

Lindamer, L. A., et al. (2014). "Gender, estrogen, and schizophrenia." Focus.

Matheson, S. L., et al. (2011). "A systematic meta-review grading the evidence for non-genetic risk factors and putative antecedents of schizophrenia." Schizophrenia research **133**(1): 133-142.

McGrath, J., et al. (2004). "A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology." BMC medicine **2**(1): 13.

Musliner, K., et al. (2015). "Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis." Psychological medicine **45**(08): 1709-1720.

NUMBER, C. O. B. O. C. and S. CASES (2010). "Rare Copy Number Variants." Arch Gen Psychiatry **67**(4): 318-327.

Nurnberger, J. I., et al. (1994). "Diagnostic interview for genetic studies: rationale, unique features, and training." Archives of general psychiatry **51**(11): 849-859.

O'donovan, M. C., et al. (2008). "Identification of loci associated with schizophrenia by genome-wide association and follow-up." Nature genetics **40**(9): 1053-1055.

Ochoa, S., et al. (2012). "Gender differences in schizophrenia and first-episode psychosis: a comprehensive literature review." Schizophrenia research and treatment **2012**.

Owen, M. J., et al. (2010). "Suggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia." Archives of general psychiatry **67**(7): 667-673.

Purcell, S., et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." The American Journal of Human Genetics **81**(3): 559-575.

Purcell, S. M., et al. (2009). "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder." Nature **460**(7256): 748-752.

Ripke, S., et al. (2013). "Genome-wide association analysis identifies 13 new risk loci for schizophrenia." Nature genetics **45**(10): 1150-1159.

Schwab, S. G. and D. B. Wildenauer (2013). "Genetics of psychiatric disorders in the GWAS era: an update on schizophrenia." European archives of psychiatry and clinical neuroscience **263**(2): 147-154.

Stone, J. L., et al. (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-241.

Sullivan, P. F., et al. (2012). "Genetic architectures of psychiatric disorders: the emerging picture and its implications." Nature Reviews Genetics **13**(8): 537-551.

Sullivan, P. F., et al. (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Archives of general psychiatry **60**(12): 1187-1192.

Appendix A: Tables and Figures

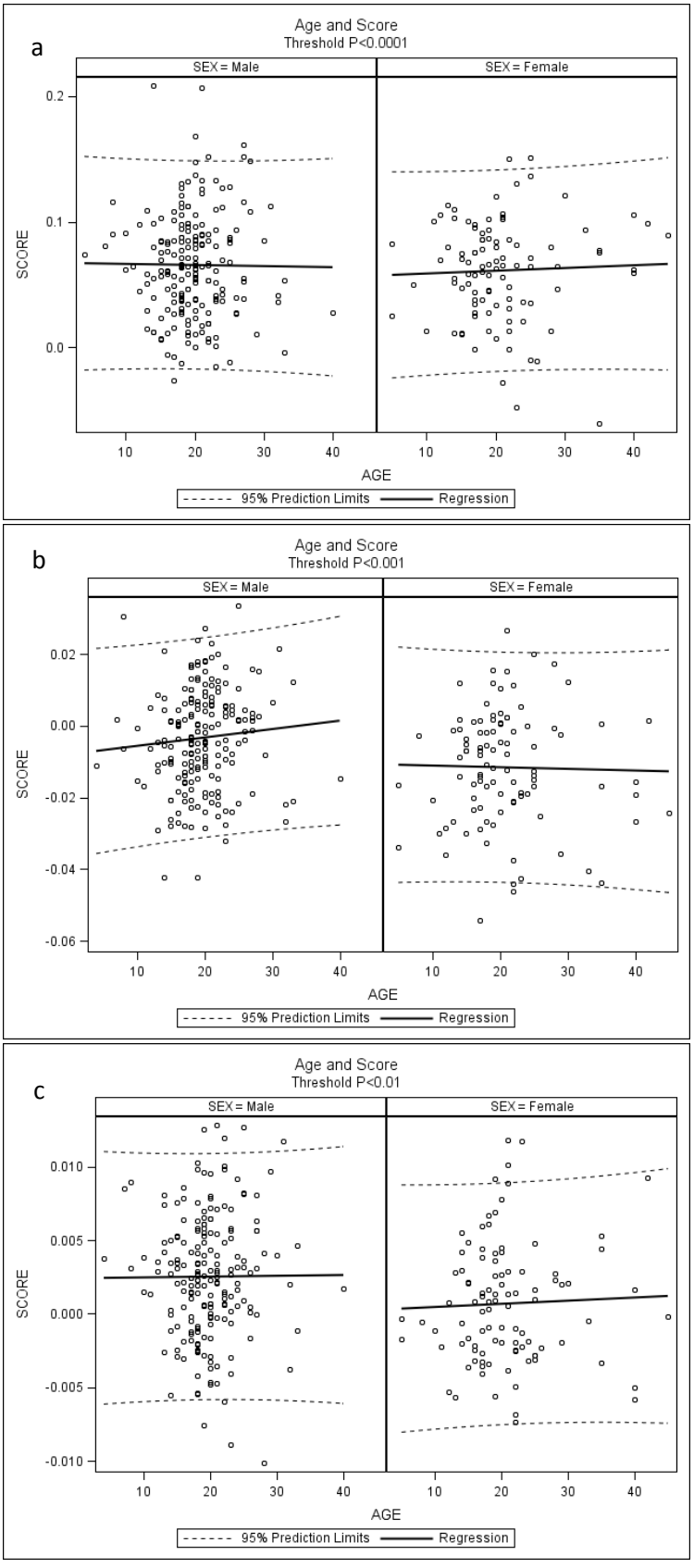

Figure 1: a: Scatterplot of score and age of onset for all target SZ cases at p-value threshold p<0.0001, b: p-value threshold p<0.001, c: p-value threshold

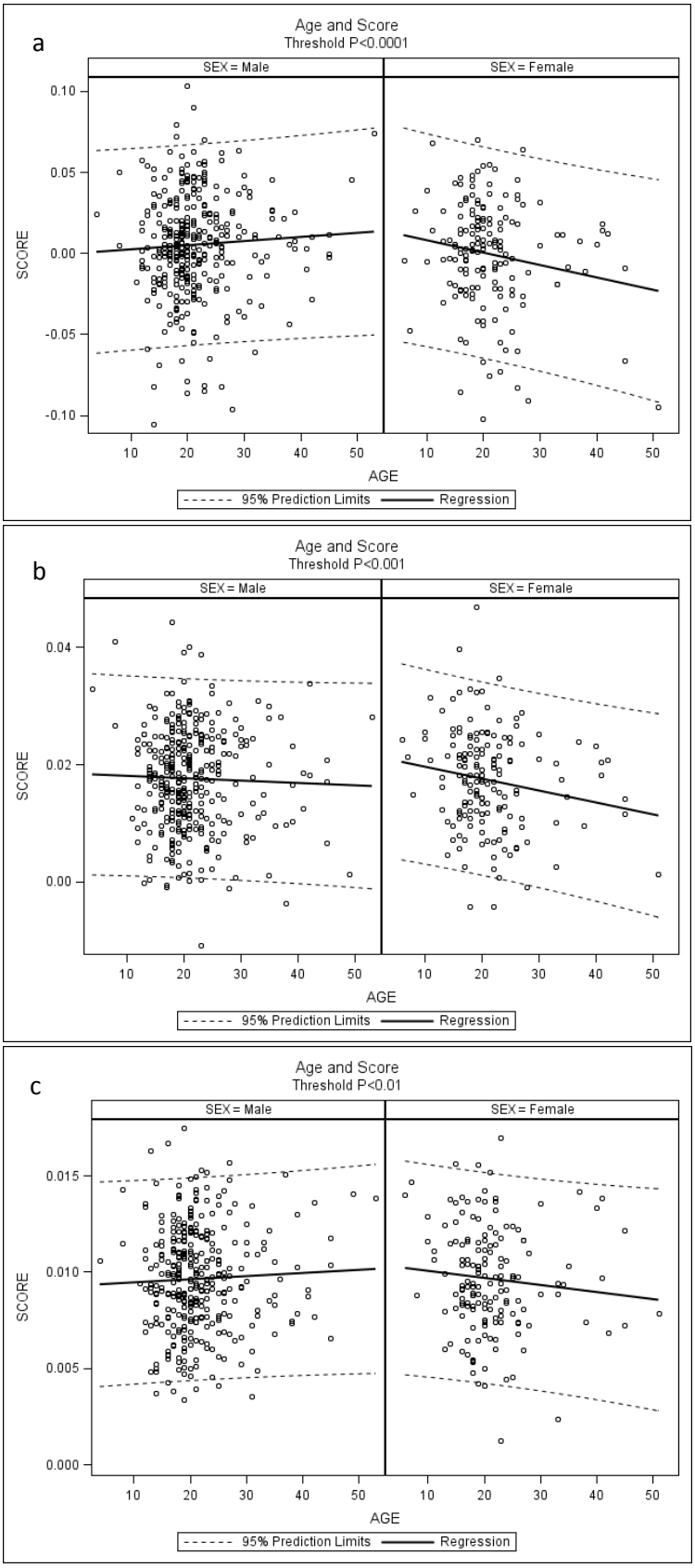

Figure 2: a: Scatterplot of score and age of onset for all SZ cases at p-value threshold p<0.0001, b: p-value threshold p<0.001, c: p-value threshold

Table 1: Demographic information[a]

	Male	Female
Cases [N(%)]	611 (58.4)	435 (41.6)
SZ[b]	401 (66.8)	199 (33.2)
BP[c]	210 (47.1)	236 (52.9)
Controls	223 (43.9)	285 (56.1)
Age of Onset [Mean (SD)]	20.8 (5.5)	22.2 (8.2)
SZ	20.1 (5.1)	20.4 (6.7)
BP	22.7 (6.3)	24.7 (9.3)
SNPs	364227	
Genotyping Rate	0.998296	

[a]Data from the Epidemiology Genetics Program at the Johns Hopkins University School of Medicine

[b]Schizophrenia

[c]Bipolar

Table 2: Comparison of Discovery and Target Datasets

	Discovery Data Set		Target Data Set	
	1	2	1	2
Cases				
SZ	296 (49.3)	306 (51.0)	304 (50.7)	294 (49.0)
BP	225 (50.4)	223 (50.0)	221 (49.6)	223 (50.0)
Controls	256 (50.4)	248 (48.8)	252 (49.6)	260 (51.2)
Sex				
Male	418 (50.1)	423 (49.0)	416 (49.9)	441 (51.0)
Female	359 (49.9)	354 (49.2)	361 (51.1)	366 (50.8)
Age of Onset	21.34 (6.54)	20.01 (5.81)	21.43 (6.67)	19.83 (9.19)

Table 3: Mean Scores of Affected and Unaffected by Dataset

Dataset		SZ scores with SZ association			BP scores with SZ association			Target Scores 1 SZ association[a]			Target Scores 2 SZ association[b]		
P-Value Threshold	Affected Status	Mean	SD	P-value	Mean	SD	P-value	Mean	SD	P-value	Mean	SD	P-value
0.0001	Unaffected	-0.0274	0.0345	<0.0001	-0.0274	0.0345	<0.0001	0.0650	0.0370	0.6058	-0.0367	0.0545	0.4903
	Affected	0.0043	0.0314		-0.0122	0.0347		0.0640	0.0410		-0.0334	0.0585	
0.001	Unaffected	-0.0058	0.0084	<0.0001	-0.0058	-0.0065	<0.0001	-0.0077	0.0163	0.2196	-0.0105	0.0139	0.1428
	Affected	0.0175	0.0086		0.0067	0.0059		-0.0060	0.0152		-0.0088	0.0137	
0.01	Unaffected	-0.0062	0.0026	<0.0001	-0.0062	-0.0064	<0.0001	0.0008	0.0040	0.0019	-0.0015	0.0039	0.0079
	Affected	0.0096	0.0027		0.0026	0.0023		0.0019	0.0042		-0.0006	0.0040	

[a]First randomization of Discovery and Target datasets

[b]Second randomization of Discovery and Target datasets

Table 4: Univariate Associations of Score and Quartile for Target Dataset

P<0.0001

Quartile	N	Affected N	Risk	95% Confidence Interval		Risk Difference	95% Confidence Interval		Risk Ratio	95% Confidence Interval		Odds Ratio	95% Confidence Interval	
1	138	73	0.5290	0.4457	0.6123	0.0000	ref		1.00	ref		1.00	ref	
2	139	68	0.4892	0.4061	0.5723	-0.0398	-0.1574	0.0779	0.92	0.73	1.17	0.85	0.53	1.37
3	139	74	0.5324	0.4494	0.6153	0.0034	-0.1142	0.1209	1.01	0.81	1.26	1.01	0.63	1.63
4	138	79	0.5725	0.4899	0.6550	0.0435	-0.0738	0.1607	1.08	0.87	1.34	1.19	0.74	1.92

P<0.001

Quartile	N	Affected N	Risk	95% CI		RD	95% CI		RR	95% CI		OR	95% CI	
1	138	63	0.4565	0.3734	0.5396	0.0000	ref		1.00	ref		1.00	ref	
2	139	79	0.5683	0.4860	0.6507	0.1118	-0.0052	0.2288	1.24	0.99	1.57	1.57	0.98	2.52
3	139	76	0.5468	0.4640	0.6295	0.0902	-0.0270	0.2075	1.20	0.95	1.52	1.44	0.89	2.30
4	138	76	0.5507	0.4677	0.6337	0.0942	-0.0232	0.2117	1.21	0.95	1.53	1.46	0.91	2.34

P<0.01

Quartile	N	Affected N	Risk	95% CI		RD	95% CI		RR	95% CI		OR	95% CI	
1	138	66	0.4783	0.3949	0.5616	0.0000	ref		1.00	ref		1.00	ref	
2	139	71	0.5108	0.4277	0.5939	0.0325	-0.0852	0.1502	1.07	0.84	1.36	1.14	0.71	1.82
3	139	70	0.5036	0.4205	0.5867	0.0253	-0.0924	0.1430	1.05	0.83	1.34	1.11	0.69	1.77
4	138	87	0.6304	0.5499	0.7110	0.1522	0.0363	0.2681	1.32	1.06	1.64	1.86	1.15	3.01

Table 5: Univariate Analysis of Phenotype

	At P<0.0001			At P<0.001			At P<0.01		
All Data	OR	95% Confidence Interval		OR	95% Confidence Interval		OR	95% Confidence Interval	
Sex	0.39	0.30	0.50						
Continuous Score	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	<0.001	>999.999
Score Quartiles	2.52	2.22	2.87	34.49	22.53	52.80	>999.999	<0.001	>999.999
Target Dataset without Bipolar									
Sex	0.42	0.29	0.59						
Continuous Score	2.83	0.15	54.01	>999.999	0.05	>999.999	>999.999	>999.999	>999.999
Score Quartiles	1.07	0.92	1.25	1.11	0.96	1.29	1.20	1.03	1.39
Bipolar									
Sex	0.88	0.68	1.14						
Continuous Score	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999
Score Quartiles	1.58	1.40	1.78	4.19	3.52	4.99	71.53	40.02	127.86

Table 6: Multivariate Analysis of Phenotype

	At P<0.0001			At P<0.001			At P<0.01		
All Data	OR	95% Confidence Interval		OR	95% Confidence Interval		OR	95% Confidence Interval	
Continuous Score and Sex	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999	>999.999
Score Quartiles Sex	2.47	2.17	2.82	33.47	21.82	51.36	>999.999	<0.001	>999.999
Target Dataset without Bipolar									
Continuous Score and Sex	2.32	0.11	47.39	>999.999	0.15	>999.999	>999.999	>999.999	>999.999
Score Quartiles Sex	1.06	0.91	1.23	1.13	0.97	1.31	1.20	1.03	1.40

Table 7: Univariate Analysis of Age

	At P<0.0001			At P<0.001			At P<0.01		
All Data	Estimate	95% CI		Estimate	95% CI		Estimate	95% CI	
Sex	-0.22	-1.50	1.06						
Continuous Score	-4.12	-22.67	14.43	-61.35	-130.88	8.18	-12.96	-233.61	207.68
Score Quartiles	-0.17	-0.75	0.40	-0.67	-1.55	0.21	-0.10	-1.05	0.84
Continuous Score by Sex									
Male	11.47	-10.95	33.90	-24.25	-106.19	57.70	101.47	-162.59	365.53
Female	-35.31	-68.58	-2.04	-146.47	-277.80	-15.14	-244.62	-648.01	158.76
Score Quartiles by Sex									
Male	0.25	-0.45	0.95	-0.28	-1.31	0.75	0.30	-0.80	1.41
Female	-1.01	-2.04	0.02	-1.55	-3.20	0.10	-1.04	-2.83	0.76
Target Dataset									
Sex	0.72	-0.78	2.23						
Continuous Score	2.90	-9.32	15.12	-28.40	-80.44	23.64	130.61	-48.01	309.23
Score Quartiles	-0.10	-0.74	0.53	-0.37	-1.02	0.29	0.32	-0.32	0.95
Continuous Score by Sex									
Male	7.42	-5.10	19.94	-13.49	-72.76	45.78	76.07	-124.49	276.64
Female	-11.66	-41.33	18.01	-52.69	-152.99	47.60	222.57	-130.51	575.65
Score Quartiles by Sex									
Male	0.34	-0.33	1.01	-0.18	-0.92	0.55	0.18	-0.54	0.89
Female	-1.19	-2.59	0.21	-0.73	-2.05	0.59	0.54	-0.71	1.78
Bipolar									
Sex	1.98	0.10	3.85						
Continuous Score	4.25	-23.80	32.31	11.01	-101.94	123.95	49.65	-304.01	403.32
Score Quartiles	0.04	-0.83	0.92	0.24	-0.83	1.32	1.02	-0.69	2.72
Continuous Score by Sex									
Male	-6.54	-36.36	23.28	-104.23	-227.31	18.86	-163.74	-552.68	225.20
Female	18.81	-29.93	67.54	143.77	-43.77	331.30	335.44	-255.49	926.37
Score Quartiles by Sex									
Male	-0.28	-1.27	0.71	-0.72	-1.97	0.52	0.16	-1.78	2.11
Female	0.36	-1.04	1.77	1.14	-0.56	2.84	1.89	-0.86	4.64

Table 8: Univariate Associations of Score and Quartile for Entire Dataset

P<0.0001

Quartile	N	Affected N	Risk	95% Confidence Interval		Risk Difference	95% Confidence Interval		Risk Ratio	95% CI		Odds Ratio	95% Confidence Interval	
1	277	46	0.2491	0.1982	0.3000	1.0000	ref		1.00	ref		1.00	ref	
2	277	114	0.4116	0.3536	0.4695	0.1625	0.0853	0.2396	1.65	1.29	2.12	2.11	1.47	3.03
3	277	186	0.6715	0.6162	0.7268	0.4224	0.3472	0.4976	2.70	2.16	3.36	6.16	4.26	8.92
4	277	231	0.8339	0.7901	0.8778	0.5848	0.5176	0.6520	3.35	2.71	4.13	15.14	9.97	22.98

P<0.001

Quartile	N	Affected N	Risk	95% CI		RD	95% CI		RR	95% CI		OR	95% CI	
1	277	2	0.0072	0.0000	0.0172	1.0000	ref		1.00	ref		1.00	ref	
2	277	69	0.2491	0.1982	0.3000	0.2419	0.1900	0.2938	34.50	8.54	139.34	45.61	11.05	188.20
3	277	252	0.9097	0.8760	0.9435	0.9025	0.8673	0.9377	126.00	31.65	501.53	1386.00	324.99	5910.92
4	277	277	1.0000	1.0000	1.0000	0.9928	0.9828	1.0000	138.50	34.81	551.01	N/A		

P<0.01

Quartile	N	Affected N	Risk	95% CI	
1	277	0	0.0000	0.0000	0.0000
2	277	46	0.1661	0.1222	0.2099
3	277	277	1.0000	1.0000	1.0000
4	277	277	1.0000	1.0000	1.0000

Table 9: Univariate Associations of Score and Quartile for Bipolar Dataset

P<0.0001

Quartile	N	Affected N	Risk	95% Confidence Interval		Risk Difference	95% Confidence Interval		Risk Ratio	95% Confidence Interval		Odds Ratio	95% Confidence Interval	
1	238	75	0.3151	0.2561	0.3741	0.0000	ref		1.00	ref		1.00	ref	
2	239	97	0.4059	0.2561	0.3741	0.0907	0.0049	0.1765	1.29	1.01	1.64	1.48	1.02	2.16
3	239	119	0.4979	0.4345	0.5613	0.1828	0.0962	0.2694	1.58	1.26	1.98	2.16	1.48	3.13
4	238	155	0.6513	0.5907	0.7118	0.3361	0.2516	0.4207	2.07	1.68	2.55	4.06	2.77	5.95

P<0.001

Quartile	N	Affected N	Risk	95% CI		RD	95% CI		RR	95% CI		OR	95% CI	
1	238	22	0.0924	0.0556	0.1292	0.0000	ref		1.00	ref		1.00	ref	
2	239	72	0.3013	0.2431	0.3594	0.2088	0.1400	0.2776	3.26	2.09	5.07	4.23	2.52	7.11
3	239	137	0.5732	0.5105	0.6359	0.4808	0.4081	0.5535	6.20	4.10	9.37	13.19	7.93	21.92
4	238	215	0.9034	0.8658	0.9409	0.8109	0.7584	0.8635	9.77	6.55	14.58	91.78	49.66	169.63

P<0.01

Quartile	N	Affected N	Risk	95% CI	
1	238	0	0.0000	0.0000	0.0000
2	239	15	0.0628	0.0320	0.0935
3	239	193	0.8075	0.7575	0.8575
4	238	238	1.0000	1.0000	1.0000