Normative Concerns Regarding Genome-Wide Association Studies and Polygenic Risk Scores for Schizophrenia

STOR 390 Midterm Project

Neel Singh

2024-10-25

Introduction

Schizophrenia is a chronic and severe mental disorder that affects how a person thinks, feels, and behaves.[1] It is characterized by symptoms such as delusions, hallucinations, disorganized thinking, and impaired functioning. These symptoms can be extremely disruptive to a person's ability to lead a normal life. The age of onset for schizophrenia is typically in early adulthood and can lead to long-term disability, making it a significant burden for individuals and their families.[1] Furthermore, individuals with schizophrenia often face stigma and discrimination, which can exacerbate their social isolation and limit opportunities in life.[2] Much of this discrimination comes from the depiction of individuals with schizophrenia as being prone to violence and criminal behavior. While most people with schizophrenia are not violent, research indicates that a small subset of individuals with schizophrenia have an increased risk of engaging in violent behavior. This violence is linked to substance abuse, a lack of access to appropriate psychiatric care, and co-occurrence during acute psychotic episodes where an individual may experience a distorted sense of reality.[3]

The exact cause of schizophrenia remains elusive. However, it is widely accepted that both biological and environmental factors contribute to its development.[1] Given the complexity of schizophrenia, understanding the full spectrum of genetic risk factors is crucial to unlocking new pathways for early detection and treatment. This is where *genome-wide association studies (GWAS)* come into play. GWAS helps to understand the genetic architecture of schizophrenia by identifying genetic variants that are associated with the disease.[4] By identifying these variants, an individual's genetic risk of developing schizophrenia can be ascertained.

From a psychiatric research perspective, understanding which genetic variations increase an individual's risk of developing schizophrenia provide insight into the biological pathways that may be disrupted in schizophrenia, offering potential targets into its etiology and potential targets for new therapies.[5] While this research holds great promise for advancing the quality of care for those with schizophrenia, it also raises ethical issues that cannot be ignored. This report examines the methodology, results, and ethical implications of an early schizophrenia GWAS titled *"Genome-wide association study identifies five new schizophrenia loci"* by Ripke and colleagues.[5]

Summary of Methods

This study involved multiple research groups pooling data from different European countries to identify specific genetic variations, or single nucleotide polymorphisms (SNPs) that are associated with schizophrenia. The study was performed in two stages:

Stage One: Data was collected from individuals with schizophrenia (cases) and compared to individuals without schizophrenia (controls) to find differences in their genetic code, or SNPs. These SNPs are then analyzed to identify likely genetic markers that are strongly associated with schizophrenia.

Stage Two: Results from the first stage were replicated and validated on a new group of individuals to confirm that the identified SNPs are consistently associated with schizophrenia.

Data Collection

For stage one, 17 datasets from 11 countries were collected, including 21,856 individuals of European ancestry. This was broken down into 9,394 cases and 12,462 controls. This ethic homogeneity was necessary to prevent confounding due to population structure, since genetic differences between ancestries can lead to false associations between SNPs and the disease of interest (schizophrenia).[4] All analyses and pre-processing steps were performed separately for each datasets.

Quality Control

Prior to analysis, the researchers performed many quality control checks on the genetic data to ensure reliability.

The researchers excluded any SNPs that were not measured in more than 5% of individuals. In other words, rare genetic variations were excluded from analysis since they are likely not generalizable to broader populations. Additionally, individuals whose genotyping was incomplete in more than 2% of SNPs were excluded. This helps ensure that individuals missing too much data did not bias the results. Then, after the first round of removing rare SNPs and individuals with many missing SNPs, the researchers removed SNPs that were not measured in more than 2% of the population. This served as an extra step to account for SNPs that became rare after excluding individuals with too many missing SNPs. Finally, SNPs that were not separately measured in 2% of cases and 2% of controls were excluded to ensure comparability between groups.

Next, SNP frequency was compared to reference values from the *HapMap (Haplotype Map) Project*. This was an international effort to map SNPs and provides a reference for comparing genetic variation.[6] Any SNP in this study whose frequency differed greater than 15% from the HapMap reference values were removed.

A statistical test to ensure *Hardy-Weinberg Equilibrium*, a principle from evolutionary biology that relates to the frequency of genotypes in large and random populations, was performed.[4] This test checked whether the SNP frequencies in the control group match what would be expected in a random sample of the population of individuals with European ancestry. SNPs that failed this test were removed as their presence could indicate errors in the genotyping or other issues with genomic data collection.

SNPs that were in *linkage disequilibrium*, or the non-random association of SNPs that are typically inherited together, were removed from the datasets.[7] This is important because if SNPs are too closely associated with one another, they may provide overlapping genetic signals. This may potentially skewing results and reducing the precision of identifying independent associations between SNPs and outcome (schizophrenia).[4] The threshold for high linkage disequilibrium was defined as a r^2, or correlation, between SNPs greater than 0.05.

After applying these checks, the datasets were narrowed down to 11,310 SNPs across individuals.

Advanced Quality Control

After ensuring quality of their datasets, the researchers began more advanced quality control measures. *Relatedness testing* is performed to identify how closely individuals in a dataset are genetically related.[8] This is important because bias can be introduced from individuals who are related to one another, since independent between samples is violated. The researchers used the *pi-hat* measure to quantify the proportion of genome shared between two individuals due to common ancestry. Pairs of individuals with a pi-hat greater than 0.9 were considered "identical samples," suggesting that they are likely duplicates. For these identical samples, SNPs that differed between each pair were randomly selected to be the "correct" SNP for the joining of identical samples. Pairs of individuals with a pi-hat greater than 0.2 but less than 0.9 were considered to be related, indicating that they are likely close family members. From families, only one randomly selected individual was kept.

After this process, the participants were reduced to 9,394 cases and 12,462 controls.

Population structure refers to the subtle genetic differences between populations that can confound GWAS results.9 To account for this, the researchers used principal component analysis (PCA). PCA is a method that reduces data complexity by find patterns across individual's genetic variation and creating new variables, or principal components that capture the most important information.[10] The researchers estimated the first 20 principal components and tested how each one was linked to schizophrenia with logistic regression. This helps elucidate if genetic patterns in each principal component are associated with the trait. Then, a method called genomic control was used to check if the results were inflated or biased in some way. Based on the *logistic regression* and genomic control tests, principal components 1, 2, 3, 4, and 6 were selected for further analysis. Logistic regression is used to model the probability of a binary outcome (in this case, having schizophrenia or not) based on one or more predictor variables (in this case, SNP genotypes).[11]

Imputation

Due to the cost of sequencing an individual's entire genome, researchers typically sequence a subset of common genetic variants and apply imputation techniques to predict the unmeasured variants.[12] Imputation relies heavily on linkage disequilibrium. In the quality control process, linkage disequilibrium is seen as a disadvantage because it leads to redundancy, or multiple SNPs that are associated with one another. During imputation, linkage disequilibrium allows researchers to fill in the gaps of unsequenced regions of individual's genomes. Here, the redundant SNPs from imputation don't interfere with the analysis since they are just filling in gaps.

The researchers drew random subsets of 300 individuals and performed imputation for each subset separately based on HapMap data.

Statistical Association Tests

After finalizing the datasets for stage one, the researchers performed logistic regression to compare the SNPs between cases and controls. For each SNP, a p-value was calculated, which indicates how likely it is that the association between a SNP and schizophrenia occurred by chance. Given the tens of thousands of SNPs tested, the researchers had to account for the fact that many associations could arise by chance simple due to the large number of tests.11 To do this, the researchers set a threshold for statistical significance far below the typical value $\alpha = 0.05$. SNPs with a p-value below a threshold of $2x10^{-}5$ were considered candidates for further investigation and validation in stage 2.

Power Analysis

A power analysis was then performed to determine how likely the study was to detect SNPs associated with schizophrenia. This depended on the *effect sizes* (odds ratios from the previous logistic regression) of those SNPs and their *allele frequencies* (how common a particular variant of a gene is within a population).[13] SNPs with more common alleles are easier to detect because they are present in more individuals, making it easier to observe their effects on schizophrenia. Rare alleles are harder to detect because fewer individuals have them, making the statistical power to detect their effects lower (especially when effect size is also small). The analysis showed that the study had a high power to detect SNPs with moderate effect sizes but lower power for those with smaller effect sizes.

Replication in Stage Two

To validate the findings of stage one, the researchers tested the SNPs identified as significant in a new group of individuals. This included 8,442 cases and 21,397 controls across 19 datasets and 14 countries. This replication step helps ensure that the results are not due to random variation or noise in the original datasets. First, the top SNPs identified in the current study and previously completed schizophrenia GWAS were validated in the stage two dataset separately. They conducted a *meta-analysis*, which combined the results from all stage two datasets together. Then, a full meta-analysis was conducted with results from all stage one and two datasets to strengthen the evidence for association between SNPs and schizophrenia.

Conditional Analysis

In some instances, more than one SNP within a region of an individual's genetic code may be associated with schizophrenia. To see if these associations are independent of one another, the researchers conducted *conditional analyses*. This is a method that checks if the association of one SNP remains significant after accounting for the effect of another SNP in the same region.[4]

Final Analyses

A combined analysis was performed after the authors realized that many SNPs associated with schizophrenia were the same as those previously known to be associated with bipolar disorder. The same methods described above were applied to 16,347 joint cases (those with schizophrenia, schizoaffective disorder, or bipolar disorder) and 14,044 joint controls.

The authors completed testing by performing a *score analysis* to test the disorders for a *polygenic model of inheritance*.[14] This means that the disorders are influenced by many different factors, each making small contributions to an individual's condition.

Summary of Results

Stage one of the study identified five new genetic regions, or *loci*, that are associated with schizophrenia, along with confirming two regions that were already known. Notably, 136 associations reached significance, with 129 mapping to the major histocompatibility complex (MHC). The MHC is a region of the genome that contains genes crucial for the immune system's ability to recognize and respond to foreign molecules. Additional novel loci identified were located at 1p21.3, 2q32.3, 8p23.2, 10q24.32 and 10q24.33. Locations like 1p21.3 refer to specific regions on the chromosome. The "1" indicates that the location of interest is on chromosome 1, where "p" indicates the short arm of the chromosome (versus the long "q" arm). The number "21" represents the region of the p arm of chromosome 1 and "3" specifies the subregion of region 21. The most significant new association was at 1p21.3. This is near a gene called MIR137, which helps control the development of brain cells. The results of this association suggest that dysregulation in MIR137 and the resulting changes in how the brain grows may play a role in schizophrenia development.

The replication phase in stage two confirmed the consistency of the initial findings, which the majority of the SNPs showing the same effect direction across the replication datasets. The study found that risk variants identified led to relatively small individual effects, with odds ratios for developing schizophrenia around 1.10 per individual risk variant. However, they collectively explain a large amount of schizophrenia's genetic risk, which supports that schizophrenia is a polygenic model.

The study also identified that certain genetic regions involving the genes CACNA1C, ANK3, and the ITIH3-ITIH4 regions are linked to both schizophrenia and bipolar disorder, suggesting that whole both disorders are distinct, they may share some underlying biological causes. This is important because it shows that psychiatric disorders can be more interconnected that previously thought.

Overall, the study expanded the known genetic architecture of schizophrenia and suggested new biological mechanisms, particularly disfunction of MIR137, in the development of the disorder. The findings also support a polygenic model of inheritance for schizophrenia, emphasizing that the cumulative effects of risk variants contribute to the total genetic risk of developing schizophrenia (although genetic risk is not necessarily the only factor in developing schizophrenia). These findings also emphasized the genetic overlap between schizophrenia and bipolar disorder, indicating common pathways in the development of both diseases.

Description of Normative Concern

The understanding of schizophrenia as a polygenic condition opens the door to the creation of *polygenic risk scores (PRS)*. PRS estimate an individual's chance of developing schizophrenia based on their genetic makeup.[4] While the ability to predict genetic risk offers the prospect of early detection and more personalized treatment, it also raises significant normative concerns that cannot be ignored.

Schizophrenia is a highly stigmatized disorder that is often misunderstood by society, with affected individuals facing prejudice in areas such as employment, insurance, and social interactions.[2] The introduction of PRS could exacerbate this discrimination. If employers, insurers, or even educational institutions gain access to genetic risk profiles, individuals with higher PRS scores for schizophrenia could be treated unfairly, even if they never develop the disorder. For example, an individual with a high genetic risk score might be denied insurance coverage or face inflated premiums based on their perceived risk of developing schizophrenia. This would unjustly penalize individuals based on genetic predispositions rather than actual health outcomes.

Although the legal protections against genetic discrimination exist, they are not comprehensive. For example, although the *Genetic Information Nondiscrimination Act (GINA)* protects individuals from genetic discrimination in health insurance and employment, it has many critical gaps.[15] These include exceptions for the military and federal government, small employers with less that 15 employees, life and disability insurance, public accommodations, and voluntary wellness programs offered by health insurers.

From a *deontological* perspective, genetic discrimination based on PRS for schizophrenia is unethical because it violates the core principle of treating individuals as ends in themselves.[16] Deontology emphasizes that all individuals possess intrinsic worth and should be treated with inherent dignity, regardless of any other factors like genetic makeup and future health risks. A person's genetic predisposition, including their PRS for schizophrenia, is beyond their control and should not be used as a basis for judgement or unequal treatment. Using such information to determine a person's social or economic opportunities, such as employment, insurance, or access to services, treats individuals not as autonomous agents, but as a means to an end.

One of the central tenets of deontology is the *categorical imperative*. There are multiple formulations of the categorical imperative. The first formulation states that individuals should "act only according to that maxim whereby you can, at the same time, will that it should become a universal law."[17] In other words, an individual's actions must conform to principles that can be universally applied. If we accept the principle that individuals can be discriminated against based on their genetic predisposition, we create a world where people are judged by factors beyond their control, diminishing their humanity. In such a world, the moral value of an individual becomes conditional on their genetic profile, not their inherent dignity as a person. This is incompatible with the duty to treat all people with respect and dignity.

Moreover, genetic discrimination violated the second formulation of the categorical imperative, which states that individuals should "act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end."[17] By discriminating based on PRS for schizophrenia, small employers or life insurers are treating individuals as a mere means to an end—primarily to reduce perceived risks or financial costs. The entity discriminating using the PRS score reduces the complexity and worth of a human being to a statistical calculation about potential future health outcomes. This instrumentalizes individuals, treating them as objects of risk management rather than independent beings with intrinsic value.

For instance, a small employer who chooses not to hire someone due to a high PRS for schizophrenia is essential deciding based not on the individual's qualifications, abilities, or character, but on the probability that they may develop a debilitating mental disorder in the future. This action reduces the person to their genetic predisposition, denying their current abilities and the contributions they could make. It fails to recognize the person's worth and autonomy—instead using them as a means to avoid perceived future costs.

The deontological perspective underscores that genetic discrimination based on PRS for schizophrenia is not only unethical, but also sets a dangerous precedent where people's opportunities and rights are contingent upon genetic factors over which they have no control. A deontologist would argue that people must be judged based on their actions and character in the present, not on risk assessments of their future mental health status. The findings of *"Genome-wide association study identifies five new schizophrenia loci"*, which identify schizophrenia as a polygenic condition, open the door to the use of PRS in estimating genetic risk.5 However, the application of these risk scores in contexts like employment, insurance, or healthcare would result in unjust treatment of individuals based on probabilistic risk assessments—factors far beyond their control. While the paper advances our understanding of schizophrenia's genetic basis, leading the way for the development of novel therapeutics and early psychiatric intervention, society must take great care to ensure that this knowledge is not misused in ways that undermine the worth and equal treatment of all individuals.

Citations

1. Jauhar S, Johnstone M, McKenna PJ. Schizophrenia. Lancet Br Ed. 2022;399(10323):473-486. doi:10.1016/S0140-6736(21)01730-X
2. Mannarini S, Taccini F, Sato I, Rossi AA. Understanding stigma toward schizophrenia. Psychiatry Res. 2022;318(Journal Article):114970-114970. doi:10.1016/j.psychres.2022.114970
3. Fazel S, Långström N, Hjern A, Grann M, Lichtenstein P. Schizophrenia, Substance Abuse, and Violent Crime. JAMA. 2009;301(19):2016-2023. doi:10.1001/jama.2009.675
4. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. Nat Rev Methods Primer. 2021;1(1). doi:10.1038/s43586-021-00056-9
5. Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011;43(10):969-976. doi:10.1038/ng.940
6. Belmont JW, Boudreau A, Leal SM, et al. A haplotype map of the human genome. Nature. 2005;437(7063):1299-1320. doi:10.1038/nature04226
7. Slatkin M. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-485. doi:10.1038/nrg2361
8. Eu-ahsunthornwattana J, Miller EN, Fakiola M, et al. Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data: e1004445. PLoS Genet. 2014;10(7). doi:10.1371/journal.pgen.1004445
9. Browning SR, Browning BL. Population Structure Can Inflate SNP-Based Heritability Estimates. Am J Hum Genet. 2011;89(1):191-193. doi:10.1016/j.ajhg.2011.05.025
10. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans R Soc Lond Ser Math Phys Eng Sci. 2016;374(2065):20150202-20150202. doi:10.1098/rsta.2015.0202
11. Ackerman A. STOR 455 Course Notes. Published online July 2024.
12. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511. doi:10.1038/nrg2796

13. Klein RJ. Power analysis for genome-wide association studies. BMC Genet. 2007;8(1):58-58. doi:10.1186/1471-2156-8-58
14. Xiang R, Kelemen M, Xu Y, et al. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. Genome Med. 2024;16(1):33-14. doi:10.1186/s13073-024-01304-9
15. American Medical Association. Genetic discrimination and the Genetic Information Nondiscrimination Act. Published online June 2013. Accessed October 24, 2024. https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/genetic-discrimination-policy-paper.pdf
16. Alexander L, Moore M. Deontological Ethics. In: Zalta EN, ed. The Stanford Encyclopedia of Philosophy. Winter 2021. Metaphysics Research Lab, Stanford University; 2021. Accessed October 24, 2024. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/
17. Kant I, Ellington JW, Kant I. Grounding for the Metaphysics of Morals ; with, On a Supposed Right to Lie Because of Philanthropic Concerns. 3rd ed. Hackett Pub. Co; 1993.