Orchid Genetic Risk Score: Schizophrenia

Orchid has developed advanced genetic risk scores (GRS) for a variety of diseases. Here we present our data on our GRS of schizophrenia.

Written by the Orchid Team

1. Schizophrenia

Schizophrenia is a severe and often debilitating long-term psychiatric condition, characterized by persistent cognitive deficits, frequent hallucinations and/or delusions, and social withdrawal [1]. The exact cause of the disease is unknown [2] but a study conducted on 31,524 twin pairs estimated the disease to be about 65.1 to 88.2% heritable [3].

2. Clinical Impact and Prevalence

About 1.5 million Americans (or 0.45% of the population) currently live with schizophrenia, and the disease is typically diagnosed in the early-mid 20s for males, and the early-mid 30s for females [4], [5]. The NIH has established three categories of symptoms for schizophrenia: psychotic symptoms, including delusions and hallucinations, negative symptoms, including reduced motivation and anhedonia, and cognitive symptoms, including problems with focusing or retaining information [4]. People with schizophrenia have an average lifespan that is 10 years shorter, mostly through higher suicide rates [6]. Like many psychiatric disorders, schizophrenia is very expensive for health systems and in 2001, was the 8th leading cause of DALY's (a measure of disability that quantifies the years of healthy life lost) for young adults (15-44 years) [7]. Treatment of schizophrenia is through antipsychotic drugs, ideally with substantial social support [1], though patients tend to have residual symptoms and deficits [8].

3. Genetic risk score (GRS)

A genetic risk score quantifies the degree to which an individual's genetics increases their likelihood of developing a specific disease. The GRS for schizophrenia includes 28,207 variants and was developed based on the variants identified in a study that analyzed genomes of 150,064 individuals of European ancestry [9]. The study included 36,989 cases (individuals with schizophrenia) and 113,075 healthy controls [9].

Number of variants in genetic risk score	**28,207**
Discovery GWAS(n=150,064)	Cases: 36,989 Controls: 113,075

Table 1: **Discovery cohort statistics**. Variants in GRS and sample number used in the schizophrenia GWAS.

4. Performant schizophrenia risk stratification

4.1. Validated using a large cohort of individuals with known disease status

Within the UK Biobank cohort, adults in the 99th percentile of genetic risk have a 1.44% absolute risk of schizophrenia, compared to 0.34% for the baseline rate. Baseline rate is the prevalence of the disease in the entire reference population.

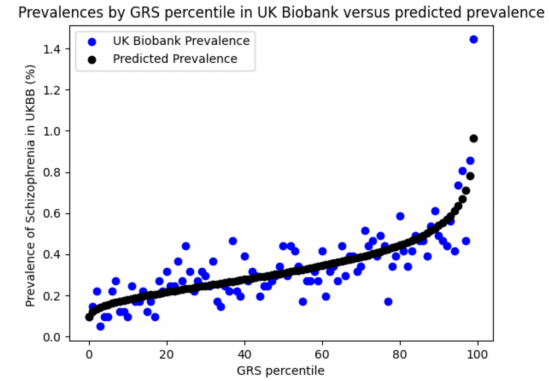

Figure 1: Risk gradient for schizophrenia. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

In the UK Biobank, cases were identified using self-reported schizophrenia (UK Biobank field 20002, code 1289) relevant ICD-9 and -10 diagnosis codes. See our supplementary table for full details. In the validation, prevalence of the disease increased with GRS. Using our phenotype definition, in our sample of self-reported white British, there were 1,386 cases of schizophrenia and 407,144 controls (prevalence of 0.34%).

Elevated Genetic Risk Definition	**Prevalence**	**Odds ratio**
Baseline Prevalence	0.34%	1
Top 5% of distribution	0.86%	2.57
Top 3% of distribution	0.92%	3.44
Top 1% of distribution	1.44%	4.34
Top 0.5% of distribution	1.71%	5.16

Table 2: Prevalence and odds ratios of schizophrenia in elevated genetic risk subgroups.

4.2. Comparing baseline and elevated risk for schizophrenia

Adults in the 99th percentile of genetic risk develop schizophrenia at 4.24 times the baseline rate, with an odds ratio of 4.34.

5. Comparison to Published Benchmarks

We compared the performance of our model as validated on the UK Biobank with the performance of a similar model in Zhang et al [10]. To make a comparison of models, we restricted our validation sample to those in Phase II of the UK Biobank release, as in Zhang et. al. In the first column, we give the results for our predictor with the phenotype as described above. In the second, we report the metrics for the best-performing predictor in Zhang et. al using the numbers reported in their paper.

Schizophrenia GRS	**Orchid**	**Reference**[1]
AUC	0.633	0.726

Table 3: **Accuracy metric comparison**. Our model compared to reference.

References

[1] Orchid Team. Overview - schizophrenia, 2022. [cited 5 Jan 2022].

[2] Orchid Team. Causes - schizophrenia., 2022. [cited 5 Jan 2022].

[3] Rikke Hilker, Dorte Helenius, Birgitte Fagerlund, Axel Skytthe, Kaare Christensen, Thomas M Werge, Merete Nordentoft, and Birte Glenthøj. Heritability of schizophrenia and schizophrenia spectrum based on the nationwide danish twin register. *Biological psychiatry*, 83(6):492–498, 2018.

[4] Orchid Team. National institute of mental health, 2022. [cited 5 Jan 2022].

[5] Orchid Team. National alliance on mental illness. mental health by the numbers., 2022. [cited 5 Jan 2022].

[6] Preben Bo Mortensen and Knud Juel. Mortality and causes of death in first admitted schizophrenic patients. *The British Journal of Psychiatry*, 163(2):183–189, 1993.

[7] Wulf Rössler, Hans Joachim Salize, Jim Van Os, and Anita Riecher-Rössler. Size of burden of schizophrenia and psychotic disorders. *European neuropsychopharmacology*, 15(4):399–409, 2005.

[8] Sonali Sarkar, Kiley Hillner, and Dawn I Velligan. Conceptualization and treatment of negative symptoms in schizophrenia. *World journal of psychiatry*, 5(4):352, 2015.

[9] Christos Pantelis, George N Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O Perkins, Olli Pietiläinen, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510):421–427, 2014.

[10] Privé F Vilhjálmsson B and Zhang Q. Improved genetic prediction of complex traits from individual-level data or summary statistics. nat commun., 2021.

6. Appendix

Disease case identification and number of cases in UK Biobank

Phenotype	**ICD-10 Codes**	**Self-Report Codes**	**Cases in UK Biobank (white British)**
Prostate cancer	C61, D075	1044	13,806
Type 2 diabetes	E11.1-9	1223	30,507
Coronary artery disease	I2104,1219,I220, I221,I228,I232 I233,I235,I236 I238,I249,I252	1075	22,451
Breast cancer	C5.0-9, D05.0, D059	1002	18,588
Inflammatory bowel disease	K51	1461,1462 ,1463	5,959
Atrial fibrillation	148.0-4,148.9	1471,1483	22,472
Schizophr enia	F20.0-9, F21, F23.0-3, F23.8	1289	1,376
Alzheimer's disease	F00.0-2, F00.9, G30.0-1,8-9.	1263	2,547
Celiac disease	K900	1456	3,253
Bipolar disease	F31	1291	1,855
Type 1 diabetes	*	*	421

Table 4: Supplementary Table: How each disease case is defined in evaluating genetic risk scores in the UK Biobank

*Type 1 diabetes was defined as a combination the following inclusion and exclusion criteria:

- Self-diagnosed diabetes (any type)
- No self-diagnosed Type 2 diabetes
- Age of diabetes onset between 0 and 20 years
- Started insulin within one year of diagnosis of diabetes