Orchid Genetic Risk Score: Atrial Fibrillation

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

Written by the Orchid Team

1. Atrial Fibrillation

Atrial fibrillation is a form of cardiac arrhythmia where the atrium of the heart beats rapidly and irregularly, which increases the risk of strokes by fivefold [1]. The precise cause of the disease is unknown [2]. The disease is substantially influenced by genetic factors, and its heritability is estimated to be approximately 60% based on a study that used data on 120,000 individuals of European ancestry [3]. Treatment of atrial fibrillation depends on a patient's risk score (CHA2DS2-VASc), which uses patient age and other risk factors to predict risk of stroke, but can include anticoagulation, slowing the heart rate, and more [4].

2. Clinical Impact and Prevalence

Atrial fibrillation is the most common form of cardiac arrhythmia; about 3-6 million Americans are believed to be living with the disease, and that figure is predicted to rise to 12 million by 2030 (3% of the population) [5]. The main complication of atrial fibrillation is the approximately 5-fold increased risk of stroke. Treatment is aimed at reducing the risk of stroke through slowing the heart rate ("rate control"), anticoagulating the blood, and occasionally catheter ablation [4].

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 atrial fibrillation includes 1,097,774 variants and was developed based on the variants identified in a study that analyzed the genomes of 133,073 individuals of European ancestry. The weights of the model were adjusted for linkage disequilibrium using PRScs software. The study included 17,931 cases (individuals with atrial fibrillation) and 115,142 healthy controls [6].

Number of variants in genetic risk score	**1,097,774**
Discovery GWAS(n=133,073)	Cases:17,931 Controls: 115,142

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

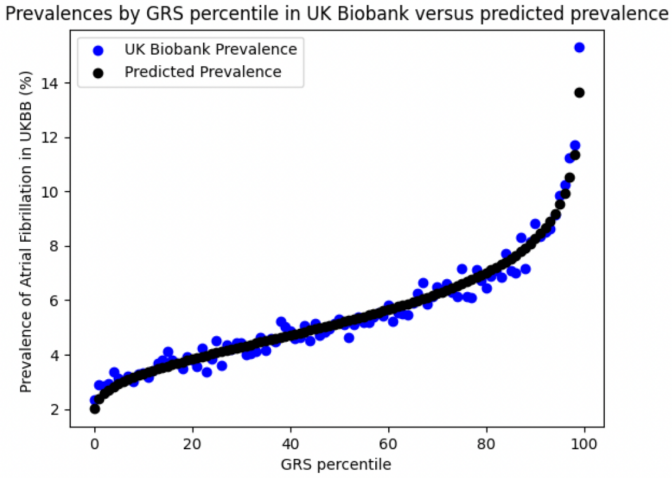

Figure 1: Risk gradient for atrial fibrillation. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

In the UK Biobank, cases were identified using self reported atrial fibrillation (UK Biobank field 20002, code) relevant ICD-10 diagnosis and death codes. See supplementary table for more details. In the validation, prevalence of the disease increased with GRS.

Prevalence at the higher end of the GRS distribution is significantly higher than the average prevalence in the UK Biobank (5.5%).

Elevated Genetic Risk Definition	**Prevalence**	**Odds ratio**
Baseline Prevalence	5.50%	1
Top 5% of distribution	11.68%	2.27
Top 3% of distribution	12.75%	2.68
Top 1% of distribution	15.30%	3.1
Top 0.5% of distribution	16.25	3.33

Table 2: **Prevalence and odds ratios in elevated genetic risk subgroups.** Adults at the tail end of GRS distribution were at an elevated risk for developing the disease in comparison to the baseline rate in the reference population, who had a prevalence of 5.5%.

3.1. Comparing Baseline and elevated risk for atrial fibrillation

Adults in the 99th percentile of genetic risk develop atrial fibrillation at 2.78 times the baseline rate, with an odds ratio of 3.1 Baseline rate is the prevalence of the disease in the entire reference population.

4. Comparison to Published Benchmarks

Orchid's model achieves comparable stratification performance with an AUC of 0.750 compared to the benchmark at 0.750.

We compared the performance of our model as validated on the UK Biobank with the performance of the best model in Khera et al. To make a comparison of models, we restricted our validation sample to those in Phase II of the UK Biobank release, as in Khera 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 Khera et. al, using the same validation setup as ours.

Model	**Orchid**	**Reference**[1]
AUC of model with GRS, age, and PCs	0.750	0.750

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

References

[1] Philip A Wolf, Robert D Abbott, and William B Kannel. Atrial fibrillation as an independent risk factor for stroke: the framingham study. *stroke*, 22(8):983–988, 1991.

[2] Orchid Team. Atrial fibrillation - causes. Available at `https://www.nhs.uk/conditions/atrial-fibrillation/causes/`, 2021. [cited 4 Jan 2022].

[3] Lu-Chen Weng, Seung Hoan Choi, Derek Klarin, J Gustav Smith, Po-Ru Loh, Mark Chaffin, Carolina Roselli, Olivia L Hulme, Kathryn L Lunetta, Josée Dupuis, et al. Heritability of atrial fibrillation. *Circulation: Cardiovascular Genetics*, 10(6):e001838, 2017.

[4] Craig T January, L Samuel Wann, Hugh Calkins, Lin Y Chen, Joaquin E Cigarroa, Joseph C Cleveland Jr, Patrick T Ellinor, Michael D Ezekowitz, Michael E Field, Karen L Furie, et al. 2019 aha/acc/hrs focused update of the 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart rhythm society in collaboration with the society of thoracic surgeons. *Circulation*, 140(2):e125–e151, 2019.

[5] Orchid Team. Atrial fibrillation. Available at `https://www.cdc.gov/heartdisease/atrial_fibrillation.htm`, 2021. [cited 4 Jan 2022].

[6] Amit V Khera, Mark Chaffin, Krishna G Aragam, Mary E Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S Lander, Steven A Lubitz, Patrick T Ellinor, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, 50(9):1219–1224, 2018.

5. 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