Orchid Genetic Risk Score: Coronary Artery Disease

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

Written by the Orchid Team

1. Coronary Artery Disease

Coronary artery disease (CAD), also called ischemic heart disease, is the reduction in blood flow (ischemia) to the heart from obstruction of the arteries [1]. CAD is caused by the progressive accumulation of plaques (termed "atherosclerosis") which narrows the arteries, and reduces blood flow, potentially causing heart pain ("stable angina") to slowly develop over time. These plaques are also prone to rupture and subsequent clotting, which in severe cases leads to myocardial infarction ("heart attacks") [2]. CAD development is influenced by environmental factors like smoking [3] and modifiable medical risk factors (hypertension, insulin resistance), but is also substantially heritable. The heritability of coronary artery disease is estimated to be between 40% and 60% based on an analysis of 20,966 twin pairs drawn from the Swedish Twin Registry [4].

2. Clinical Impact and Prevalence

In the United States, coronary artery disease affects roughly 18.2 million adults above the age of 20, or about 6.7% [1]. The mean age of diagnosis is 57.4 for men, and 59.3 for adults [5]. Typical symptoms of the disease include chest pain ("angina") and a shortness of breath that is caused by the inability of the arteries to deliver oxygenated blood to the body [6]. In some cases, this can result in a myocardial infarction (a heart attack). CAD has no single cure, but can be managed; cardiologists prescribe combination treatments of medications such as beta blockers, statins, ACE inhibitors, and more, along with lifestyle modifications (such as smoking cessation), which reduce the risk of future heart attacks and slow atherosclerosis [7].

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 CAD includes 1,107,562 variants and was developed based on the variants identified in a study that analyzed genomes of 184,305 individuals of European ancestry [8]. The study included 60,801 cases (individuals with CAD) and 123,504 healthy controls [8]. The weights of the model were adjusted for linkage disequilibrium using PRScs software [9].

Number of variants in genetic risk score	**1,107,562**
Discovery GWAS(n=184,305)	Cases: 60,801 Controls: 123,504

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

4. Comparison to Published Benchmarks

4.1. Validated using a large cohort of real world adults with known coronary artery disease status

Over the course of the UK Biobank observation period, with a median age of 58, the adults in the 99th percentile of genetic risk had a 14% absolute risk of CAD, compared to 5.5% for the baseline rate. Importantly, this is different from the lifetime risk of heart disease in the US, which is around 40% [10] for two reasons: the UK Biobank cohort has a median age of 58 [11], so most people who will develop heart disease have not yet done so; also, the UK Biobank population is somewhat healthier than average [12]. Both of those will lower the prevalence of heart disease in the UK Biobank cohort.

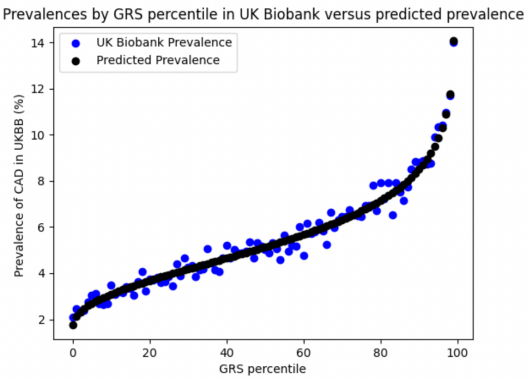

Figure 1: Risk gradient for coronary artery disease. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

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

4.2. Comparing Baseline and elevated risk for coronary artery disease

Adults in the 99th percentile of genetic risk develop coronary artery disease at 2.55 times the baseline rate, or a 2.80 odds ratio, over the observation period. Baseline rate is the prevalence of the disease in the entire reference population.

Elevated Genetic Risk Definition	Absolute Risk Over observation period	Odds ratio over observation period
Baseline Prevalence	5.50%	1
Top 5% of distribution	11.48%	2.23
Top 3% of distribution	12.21%	2.39
Top 1% of distribution	14.00%	2.8
Top 0.5% of distribution	14.29%	2.87

Table 2: Absolute risk and odds ratios over observation period in elevated genetic risk subgroups in white British individuals.

Orchid's model achieves a stratification performance with an AUC of 0.752, which is the same as the benchmark AUC of 0.752.

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 phenotype as ours.

Model	Orchid	Reference[1]
AUC of model with GRS, age, and PCs[2]	0.752	0.752

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

References

[1] Orchid Team. Heart disease facts, 2021. [27 Sep 2021].

[2] John A Ambrose and Manmeet Singh. Pathophysiology of coronary artery disease leading to acute coronary syndromes. *F1000prime reports*, 7, 2015.

[3] Ira S Ockene and Nancy Houston Miller. Cigarette smoking, cardiovascular disease, and stroke: a statement for healthcare professionals from the american heart association. *Circulation*, 96(9):3243–3247, 1997.

[4] Ruth McPherson and Anne Tybjaerg-Hansen. Genetics of coronary artery disease. *Circulation research*, 118(4):564–578, 2016.

[5] Victor Okunrintemi, Martin Tibuakuu, Salim S Virani, Laurence S Sperling, Annabelle Santos Volgman, Martha Gulati, Leslie Cho, Thorsten M Leucker, Roger S Blumenthal, and Erin D Michos. Sex differences in the age of diagnosis for cardiovascular disease and its risk factors among us adults: trends from 2008 to 2017, the medical expenditure panel survey. *Journal of the American Heart Association*, 9(24):e018764, 2020.

[6] Orchid Team. Coronary artery disease, 2021. [4 Jan 2022].

[7] Donna K Arnett, Roger S Blumenthal, Michelle A Albert, Andrew B Buroker, Zachary D Goldberger, Ellen J Hahn, Cheryl Dennison Himmelfarb, Amit Khera, Donald Lloyd-Jones, J William McEvoy, et al. 2019 acc/aha guideline on the primary prevention of cardiovascular disease: a report of the american college of cardiology/american heart association task force on clinical practice guidelines. *Circulation*, 140(11):e596–e646, 2019.

[8] 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.

[9] Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W Smoller. Polygenic prediction via bayesian regression and continuous shrinkage priors. *Nature communications*, 10(1):1776, 2019.

[10] Donald M Lloyd-Jones, Martin G Larson, Alexa Beiser, and Daniel Levy. Lifetime risk of developing coronary heart disease. *The Lancet*, 353(9147):89–92, 1999.

[11] Orchid Team. Data-field 21022., 2021. [cited 22 Feb 2022].

[12] Anna Fry, Thomas J Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E Allen. Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. *American journal of epidemiology*, 186(9):1026–1034, 2017.

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