Orchid Genetic Risk Score: Alzheimer's Disease

Written by the Orchid Team

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

1. Alzheimer's Disease

Alzheimer's disease (AD) is a chronic, irreversible, neurodegenerative disease with no known cure. It causes brain cells to die and produces harmful deposits, called amyloid plaques, which appear around the neurons of the brain [1]. There is some debate around the precise mechanism for developing Alzheimer's disease. The disease is substantially influenced by genetic factors, with its heritability being estimated between 60 and 80%, based on an analysis conducted using data on almost 12,000 twin pairs drawn from the Swedish twin registry [2], [3].

2. Clinical Impact and Prevalence

Most diagnoses of Alzheimer's disease (AD) are made in the developed world [2]. In 2020 about 5.8 million Americans (1.7%) had AD, and most were diagnosed after the age of 60 [4] with the average age of diagnosis being 75 [5]. Aside from the neurological symptoms such as cerebral atrophy and plaque deposits, AD produces a range of cognitive symptoms such as impaired memory and thinking, sleeplessness, social withdrawal, and distrust in others [6]. The disease is fatal, and although most sufferers do not live beyond 10 years after being diagnosed [7], the FDA has approved one medication, aducanumab, for treatment [1].

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 **AD** includes **1,111,805** variants and was developed based on the variants identified in a study that analyzed genomes of about **455,258** individuals of European ancestry affected by AD [8]. The study included **17008** cases (individuals with **AD**) and **37154** healthy controls.

Our AD GRS has some special characteristics relative to our usual GRS. For AD, one specific loci, APOE, confers a disproportionate share of genetic risk. That is, AD follows an oligogenic model, not a classic polygenic model. For that reason, the resulting GRS is not normally distributed.

Number of variants in genetic risk score	**1,111,805**
Discovery GWAS(n=455,258)	Cases: 71,880 Controls: 383,378

Table 1: Discovery cohort statistics.

4. Performant Alzheimer's disease risk stratification

4.1. Validated using a large cohort of real world individuals with known Alzheimer's disease status

Because the discovery GWAS data included samples from the UK Biobank, to validate we tested the performance of the GRS on the set of samples not included in that study: participants who did not have a known parental status of Alzheimer's disease. Individuals in the 99th percentile of genetic risk have a 9.90% percent prevalence of Alzheimer's disease (AD), compared to the average of 0.93% percent. This is lower than the lifetime prevalence of AD reported for US individuals, as a result of the median age of the UK Biobank (58), which means that many individuals who will eventually develop AD have not yet done so. This biases the prevalence of AD within the UK Biobank downwards substantially relative to the average lifetime prevalence of 1.7% in the US.

Validation in UK Biobank. In the UK Biobank, cases were identified using self-reported Alzheimer's disease (UK Biobank field 20002) relevant ICD-9/ICD-10 diagnosis. Because the discovery GWAS data included samples from the UK Biobank, to validate we tested the performance of the GRS on the set of samples not included in that study: participants who did not have a known parental status of Alzheimer's disease. See our supplementary table for full details. In the validation, the prevalence of Alzheimer's disease increased with GRS. We restricted our analysis to self-reported British white individuals whose genetic ancestry matched their self-identification. With our phenotype definition there were 452 cases of Alzheimer's disease and 45,909 controls.

4.2. Comparing baseline and elevated risk for Alzheimer's disease

Individuals in the 99th percentile of genetic risk develop Alzheimer's disease at 10.6 times the baseline rate. The odds ratio for individuals in the 99th percentile was 11.7. Baseline rate is the prevalence of the disease in the entire reference population.

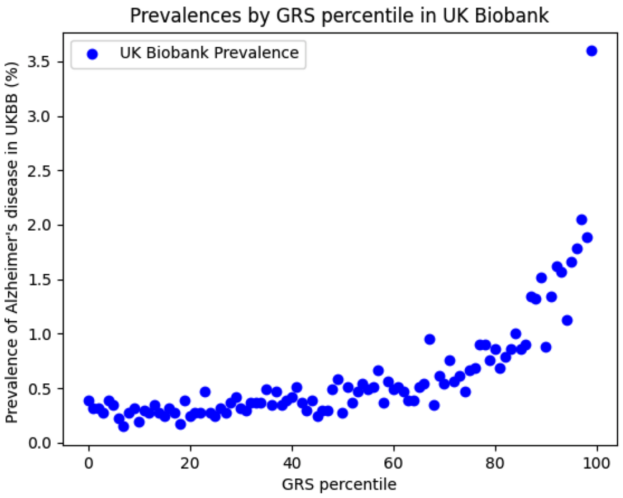

Figure 1: Risk gradient for type 1 diabetes. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis.

Elevated Genetic Risk Definition	Prevalence	Odds ratio
Baseline Prevalence	0.93%	1.00
Top 5% of distribution	5.18%	5.80
Top 3% of distribution	6.47%	7.35
Top 1% of distribution	9.91%	11.69

Table 2: Disease prevalence in elevated genetic risk subgroups for White British individuals.

5. Comparison to Published Benchmarks

Orchid's model achieves an AUC of 0.839 compared to the benchmark of 0.782.

We compared the performance of our model as validated on the UK Biobank with the performance of the best model in [8]. The benchmarks in this table were generated on different datasets, so they are not precisely comparable.

Model	Orchid	Reference
AUC of model with GRS, age, and PCs[1]	0.839	0.782

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

References

[1] Orchid Team. Alzheimer's disease fact sheet. Available at `http://www.nia.nih.gov/health/alzheimers-disease-fact-sheet`, 2022. [cited 4 Jan 2022].

[2] Caroline Van Cauwenberghe, Christine Van Broeckhoven, and Kristel Sleegers. The genetic landscape of alzheimer disease: clinical implications and perspectives. *Genetics in Medicine*, 18(5):421–430, 2016.

[3] Margaret Gatz, Chandra A Reynolds, Laura Fratiglioni, Boo Johansson, James A Mortimer, Stig Berg, Amy Fiske, and Nancy L Pedersen. Role of genes and environments for explaining alzheimer disease. *Archives of general psychiatry*, 63(2):168–174, 2006.

[4] Orchid Team. What is alzheimer's disease? Available at `https://www.cdc.gov/aging/aginginfo/alzheimers.htm`, April 2021. [cited 4 Jan 2022].

[5] Josephine Barnes, Bradford C Dickerson, Chris Frost, Lize C Jiskoot, David Wolk, and Wiesje M van der Flier. Alzheimer's disease first symptoms are age dependent: evidence from the nacc dataset. *Alzheimer's & dementia*, 11(11):1349–1357, 2015.

[6] Orchid Team. Alzheimer's disease. Available at `https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447`, 2022. [cited 4 Jan 2022].

[7] Stephen Todd, Stephen Barr, Mark Roberts, and A Peter Passmore. Survival in dementia and predictors of mortality: a review. *International journal of geriatric psychiatry*, 28(11):1109–1124, 2013.

[8] Iris E Jansen, Jeanne E Savage, Kyoko Watanabe, Julien Bryois, Dylan M Williams, Stacy Steinberg, Julia Sealock, Ida K Karlsson, Sara Hägg, Lavinia Athanasiu, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nature genetics*, 51(3):404–413, 2019.

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