Validating Orchid's Alzheimer's Disease Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Alzheimer's disease is a chronic, progressive neurodegenerative disorder that primarily affects memory, thinking, and behavior, eventually impairing a person's ability to perform everyday activities and leading to death over time.[1] The estimated lifetime prevalence of Alzheimer's disease is approximately 9.5%.[2] It is one of the leading causes of death among older adults and is associated with increased morbidity, long-term care needs, and significant health care and caregiver burden.[3] There is currently no cure for Alzheimer's disease, and the condition is associated with a progressive decline in cognitive and functional abilities, leading to increasing dependence, reduced quality of life, and premature mortality.[4] Available pharmacologic and non-pharmacologic treatments may help alleviate symptoms or modestly slow disease progression but do not halt or reverse the underlying neurodegenerative process.[5]

Genetic Risk Score

A person's risk of developing Alzheimer's disease is shaped substantially by genetics.[6] Genetic risk scores (GRS) allow us to estimate this disease risk based on the DNA of a person or embryo.[7] Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease compared to the population baseline risk.

The Alzheimer's disease GRS included in Orchid's reports has two components. The first component of the GRS was developed based on a study that analyzed the genomes of individuals of primarily European ancestry, including 17,008 cases (individuals with Alzheimer's disease) and 37,154 healthy controls.[8] This component of the GRS considers over a million variants in the DNA, each of which is thought to be associated with a very small impact on Alzheimer's risk.

The second component of our GRS for Alzheimer's disease considers one gene in particular, called apolipoprotein E (APOE), which plays a large role in shaping Alzheimer's risk. Everyone inherits one copy of the APOE gene from each parent, and there are three common versions (also called "alleles") of the gene: E2, E3, and E4. The E3 allele is the most common, and having two E3 copies is typically considered to be the baseline for Alzheimer's risk. The E4 allele increases a person's likelihood of developing Alzheimer's disease, while the E2 allele is protective.[9] In order to estimate Alzheimer's risk, we consider both the large effect on risk from APOE, as well as the small effects from over a million other variants in the DNA.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's Alzheimer's disease GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.[10] Because the discovery GWAS data included samples from the UKB, we validated the GRS by testing its performance on the set of samples not included in that study: participants whose parental status of Alzheimer's disease was unknown. Additionally, we restricted the analysis to individuals of British ancestry who were at least 55 years old. We also estimated parameters for the APOE component of the score based on this set of samples. We defined Alzheimer's disease using the G30 ICD-10 code, yielding 330 cases and 28,276 controls (approximately 1.2% observed prevalence). We then grouped individuals by GRS percentile and compared the observed disease prevalence within each group to our model's predictions (Figure 1). For additional technical details, see the Supplementary Information.

Table 1 shows the Alzheimer's disease observed prevalence for individuals in the UKB grouped by GRS percentile range (top 10%, 5%, and 1%), as well as how their risk compares to the baseline risk at the 50th GRS percentile. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of Alzheimer's disease, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

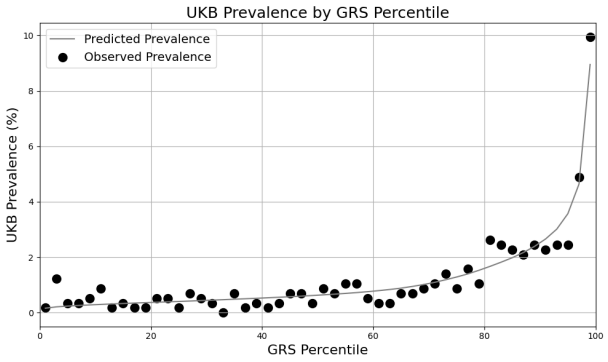

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	0.70%	1.00
Top 10%	4.40%	6.54
Top 5%	6.50%	9.87
Top 1%	13.24%	21.67

Table 1. Observed prevalence of Alzheimer's disease in the UKB by GRS percentile range. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of Alzheimer's disease.

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	6.02%	1.00x
95th	27.08%	4.50x
97th	32.55%	5.41x
99th	46.58%	7.74x

Table 2. Predicted lifetime prevalence of Alzheimer's disease at different GRS percentiles. Individuals with the highest GRS percentiles are predicted to have an increased risk of Alzheimer's disease relative to those at the 50th percentile.

Estimating Lifetime Risk

The average observed prevalence of Alzheimer's disease in the UKB was 1.2%. This is considerably lower than the lifetime prevalence, which has been estimated to be approximately 9.5%.[2] This is likely due in part to the fact that UKB participants tend to be healthier than the general population, which leads to lower observed disease prevalence.[11] Additionally, the observed prevalence in the UKB includes people still living who could develop the disease when they are older, and so does not capture the full lifetime risk of the disease.

Orchid's clinical reports include predicted lifetime disease risk, which we calculate by first estimating how disease risk varies across GRS in the UKB and then rescaling that pattern so the average matches the known lifetime population risk (Figure 2).[7] People at the high end of the GRS distribution are predicted to have an elevated lifetime risk of the disease relative to the population (Table 2).

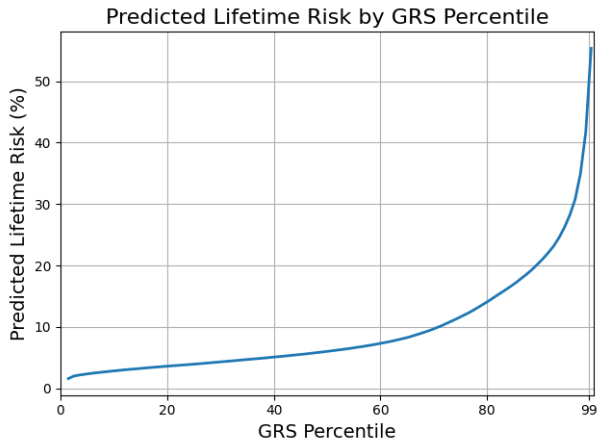

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the approximately 9.5% estimate.[2]

Conclusion

In this study, we evaluated our Alzheimer's disease GRS on data from the UKB. We found that it performed well, particularly for identifying individuals with elevated risk of the disease relative to the population. In our embryo and couple reports, we adjust the model to predict lifetime risk, which is generally higher than observed prevalence in the UKB. The Alzheimer's GRS model is available to individuals of all ancestry groups.

Acknowledgements

This research was conducted using the UK Biobank Resource under Application Number 80545.

References

1. Harvard Health Publishing. Alzheimer's Disease, April 2023. Accessed 2026-02-17. URL: https://www.health.harvard.edu/topics/alzheimers-disease.
2. Ron Brookmeyer, Denis A. Evans, Liesi Hebert, et al. National Estimates of the Prevalence of Alzheimer's Disease in the United States. *Alzheimer's & Dementia*, 7(1):61–73, 2011. doi:10.1016/j.jalz.2010.11.007.
3. Alzheimer's Association. 2024 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 20(5):3708–3821, 2024. doi:10.1002/alz.13809.
4. Mayo Clinic Staff. Alzheimer's Disease, 2024. Accessed 2026-02-17. URL: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease.
5. Alzheimer's Association. Treatments for Alzheimer's & Dementia, 2025. Accessed 2026-02-17. URL: https://www.alz.org/alzheimers-dementia/treatments.
6. C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers. The Genetic Landscape of Alzheimer Disease: Clinical Implications and Perspectives. *Genetics in Medicine*, 18(5):421–430, 2016. doi:10.1038/gim.2015.117.
7. Nilanjan Chatterjee, Jianxin Shi, Montserrat García-Closas, et al. Developing and Evaluating Polygenic

Risk Prediction Models for Stratified Disease Prevention. *Nature Reviews Genetics*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.

8. I. E. Jansen, J. E. Savage, K. Watanabe, et al. Genome-wide Meta-analysis Identifies New Loci and Functional Pathways Influencing Alzheimer's Disease Risk. *Nature Genetics*, 51(3):404–413, 2019. doi:10.1038/s41588-018-0311-9.

9. Mayo Clinic Staff. Alzheimer's Genes: Are You at Risk?, 2024. Accessed 2026-02-17. URL: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-genes/art-20046552.

10. C. Sudlow, J. Gallacher, N. Allen, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015. doi:10.1371/journal.pmed.1001779.

11. A. Fry, T.J. Littlejohns, C. Sudlow, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, 186:1026–1034, 2017. doi:10.1093/aje/kwx246.

12. L. A. Farrer, L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, R. H. Myers, M. A. Pericak-Vance, N. Risch, and C. M. van Duijn. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA*, 278(16):1349–1356, 1997. doi:10.1001/jama.1997.03550160069041.

Supplementary Information

As we noted in the Genetic Risk Score section, a substantial component of the genetic risk for Alzheimer's disease is due to specific variants in the APOE gene. Table 3 shows the Alzheimer's disease prevalence for individuals in the UKB grouped by APOE genotype. For reference, we also report literature-based OR estimates from the large meta-analysis by Farrer et al.[12] Please note that ORs in Table 3 are relative to the average E3/E3 baseline, not the 50th percentile baseline, in order to match the Farrer et al. OR reporting convention. As expected, individuals carrying at least one E4 allele show elevated Alzheimer's disease prevalence relative to the E3/E3 baseline, with E4/E4 carriers exhibiting the highest risk. Conversely, individuals carrying at least one E2 allele show reduced prevalence compared to E3/E3.

Genotype	N	N (%)	Prev. (%)	OR vs E3/E3	OR vs E3/E3 (Farrer et al.[12])
E2/E2	171	0.6%	0.00	0.00	0.6
E2/E3	3,603	12.6%	0.53	0.92	0.6
E2/E4	701	2.5%	1.57	2.76	2.6
E3/E3	16,706	58.4%	0.57	1.00	1.0
E3/E4	6,753	23.6%	2.24	3.96	3.2
E4/E4	672	2.3%	7.89	14.81	14.9

Table 3. Observed prevalence in the UKB by APOE genotype. Individuals with at least one E4 allele show elevated Alzheimer's disease risk relative to the E3/E3 baseline, with E4/E4 carriers exhibiting the highest risk, while the E2 allele is protective.

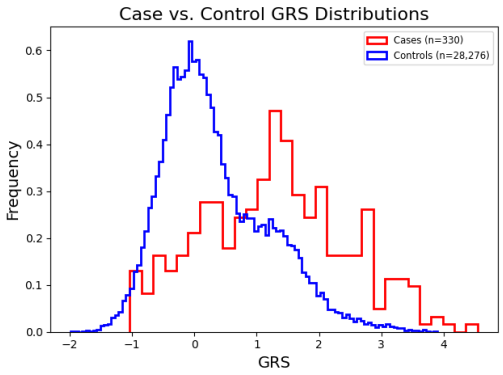

Figure 3. GRS histograms. GRS distributions for cases and controls. The case distribution is shifted noticeably higher compared to the controls.

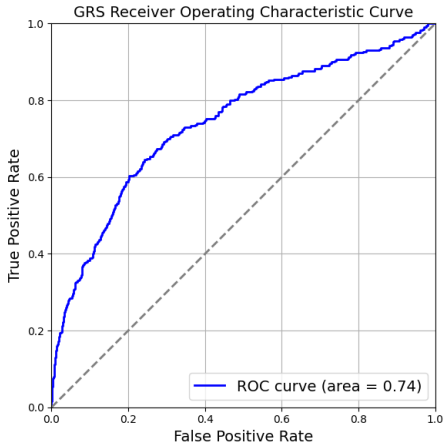

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds. A curve closer to the top-left indicates a better model, while a diagonal line (AUC = 0.5) represents random guessing.