Validating Orchid's Atrial Fibrillation Genetic Risk Score

Written by Orchid Bioinformatics Staff

ORCHID

Introduction

Atrial fibrillation is a common cardiac arrhythmia in which the atria beat rapidly and irregularly. It can cause palpitations, dizziness, shortness of breath, and fatigue. Age is the primary risk factor, but hypertension, diabetes, and obesity also increase susceptibility. It can be triggered by binge drinking or elevated stress. Atrial fibrillation can lead to serious complications, including a 5x risk of stroke[1,2].

Atrial fibrillation is the most common form of cardiac arrhythmia and is projected to affect up to 12.1 million Americans by 2030[3]. A family history of atrial fibrillation is associated with a 40% increased risk[2]. It is estimated that 37.1% of individuals over the age of 55 will develop atrial fibrillation during their lifetime[4]. Treatment is complex and may include medications to slow heart rate, prevent arrhythmia, and prevent stroke complications. Surgical options are rare and include catheter ablation and pacemaker implantation[2,5].

Genetic Risk Score

Atrial fibrillation is shaped by both environmental and genetic factors. Monogenic testing is not available because no single gene causes the condition. Genetic risk scores (GRS), which combine the small effects of many variants into a single score, are currently the only way to estimate genetic risk. Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease.

Orchid's atrial fibrillation GRS was trained following current industry standards[6,7]. The GRS was constructed using the SBayesRC algorithm trained on publicly available FinnGen and Million Veterans Program summary statistics[8,9]. The summary statistics include 151,410 cases and 778,364 controls[10]. The resulting GRS contains over one million variants.

Risk predictions are adjusted to each individual's ancestry, with predictive power decaying as genetic distance from the predominately European training data increases[11]. Orchid considers a GRS meaningfully predictive if individuals at roughly the 97.7th percentile have an odds ratio (OR) of at least 2. The atrial fibrillation GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive performance of Orchid's atrial fibrillation GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom[12]. We restricted the analysis to participants of British ancestry aged 55 or older and defined atrial fibrillation as any diagnoses under ICD-10 codes I48.x, yielding 26,093 cases and 233,099 controls (10.1% 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.
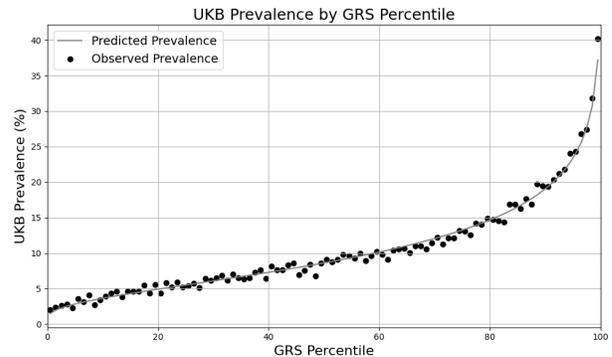

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

Table 1 shows the observed prevalence of atrial fibrillation 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 atrial fibrillation, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	8.84%	1.00
Top 10%	24.80%	3.40
Top 5%	29.30%	4.27
Top 1%	40.21%	6.93

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

Estimating Lifetime Risk

The average observed prevalence of atrial fibrillation in the UKB was 10.1%. This is considerably lower than the

lifetime prevalence in the general population, which has been estimated to be approximately 37.1%.[4] 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.[13] 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).[14] 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).
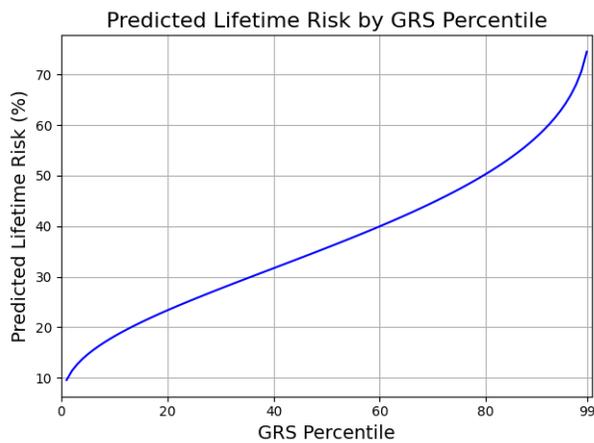

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 37.1% estimate.[4]

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	35.72%	1.00x
95th	64.26%	1.80x
97th	68.02%	1.90x
99th	74.51%	2.09x

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

Conclusion

In this study, we evaluated our atrial fibrillation 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 atrial fibrillation GRS model is available to individuals of all ancestry groups.

Acknowledgments

References

1. P. A. Wolf, R. D. Abbott, and W. B. Kannel. Atrial Fibrillation as an independent risk factor for stroke: The Framingham Study. *Stroke*, 22(8):983–988, 1991. doi: 10.1161/01.STR.22.8.983.

2. B. J. J. M. Brundel, X. Ai, M. T. Hills, et al. Atrial Fibrillation. *Nature Reviews Disease Primers*, 8:21, 2022. doi:10.1038/s41572-022-00347-9.

3. CDC. Atrial Fibrillation. https://www.cdc.gov/heart-disease/about/atrial-fibrillation.html, 2024. [cited 5 Dec 2025].

4. L. C. Weng, S. R. Preis, O. L. Hulme, et al. Genetic predisposition, clinical risk-factor burden, and lifetime risk of Atrial Fibrillation. *Circulation*, 137:1027–1038, 2018. doi:10.1161/CIRCULATIONAHA.117.031431.

5. C. T. January, L. S. Wann, H. Calkins, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with Atrial Fibrillation. *Circulation*, 140:e125–e151, 2019. doi:10.1161/CIR.0000000000000665.

6. S. Moore, I. Davidson, J. Anomaly, et al. Development and validation of polygenic scores for within-family prediction of disease risks. *medRxiv*, 2025. doi:10.1101/2025.08.06.25333145.

7. S. Cordogan, D. B. Starr, N. R. Treff, et al. Within- and between-family validation of nine polygenic risk scores developed in 1.5 million individuals: Implications for IVF, embryo selection, and reduction in lifetime disease risk. *medRxiv*, 2025. doi:10.1101/2025.10.24.25338613.

8. Z. Zheng, S. Liu, J. Sidorenko, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nature Genetics*, 56:767–777, 2024. doi:10.1038/s41588-024-01704-y.

9. FinnGen. FinnGen+MVP+UKBB Summary Statistics. https://mvp-ukbb.finngen.fi/about. Accessed: 2025-12-05.

10. FinnGen. FinnGen+MVP+UKBB Phenotypes. https://mvp-ukbb.finngen.fi. Accessed: 2025-12-05.

11. Florian Privé et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *American Journal of Human Genetics*, 109(1):12–23, 2022. doi:10.1016/j.ajhg.2021.11.008.

12. 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.

13. 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.

14. Nilanjan Chatterjee, Jianxin Shi, and Montserrat García-Closas. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.

Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
35.72%	2.04	4.17

Table 3. OR per SD. The baseline risk for an individual with a median GRS, and the predicted OR at one and two SDs, respectively. A GRS must have a predicted OR >2 at 2 SD to be included in Orchid's clinical reports.

UKB Prevalence	Population Prevalence	Liability R^2
10.1%	37.1%	13.73%

Table 4. Liability R^2. The estimated liability R^2 using a population prevalence of 37.1%.
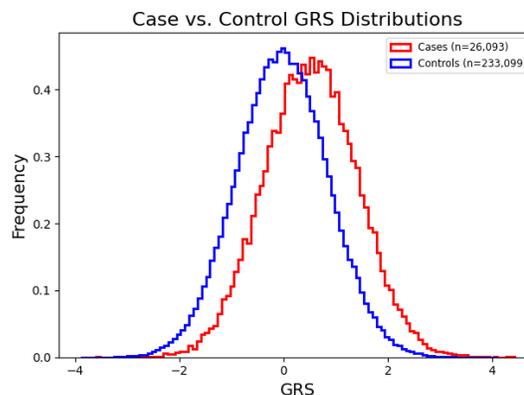

Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.
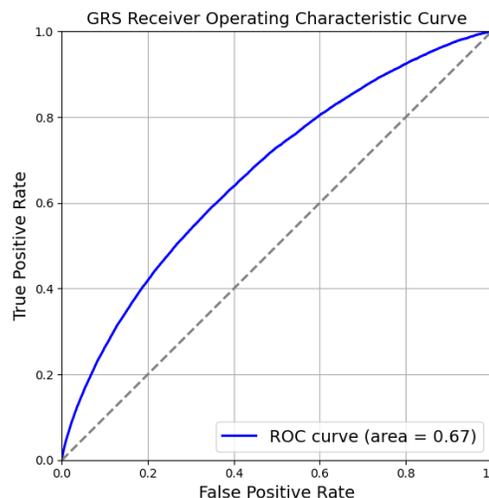

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds. A curve closer to the top-left indicates a better model, while a diagonal line (AUC = 0.5) represents random guessing.
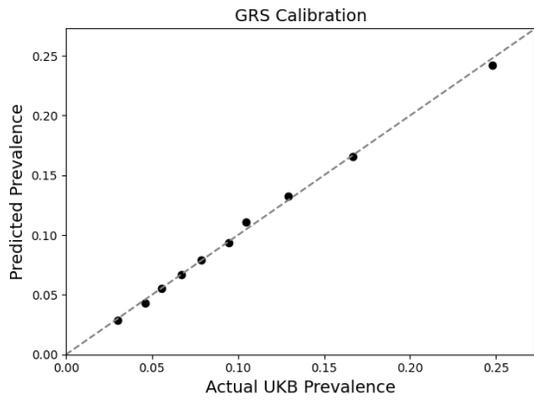
Figure 5. **Calibration Curve.** Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.