Validating Orchid's Breast Cancer Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Breast cancer is a disease in which malignant cells form in breast tissue. It most commonly presents with physical changes to the breast, such as a new lump or thickening, changes in breast size or shape, or skin changes involving the breast or nipple.[1]

Risk for breast cancer is influenced by both genetic and non-genetic factors. Non-genetic risk factors include increasing age, reproductive history such as having a first menstrual period before age 12 or menopause after age 55, use of certain hormone therapies, alcohol consumption, being overweight or obese after menopause, lack of regular physical activity, and exposure to diethylstilbestrol (DES).[2] A small proportion of cases are associated with inherited high-risk genetic variants, including pathogenic variants in the BRCA1 and BRCA2 genes, which substantially increase lifetime risk of breast cancer.[3,4]

There are currently more than 4 million breast cancer survivors in the United States, and approximately 13% of women in the US will be diagnosed with the disease in their lifetime.[5,6] Prognosis for breast cancer varies considerably by stage at diagnosis. The 5-year relative survival rate is more than 99% for localized disease (cancer confined to the breast), about 87% for regional disease, and about 32% for metastatic disease.[7] Mammography screening can help with earlier detection, and several treatments such as medications (including chemotherapy), surgery, and radiotherapy may be prescribed by an oncologist.[8]

Genetic Risk Score

Breast cancer is shaped by both environmental and genetic factors. Some rare monogenic variants such as BRCA1/2 are known to substantially increase breast cancer risk[3,4] but most cases arise from the combined effects of many genetic variants and environmental exposures. A minority of individuals carry a pathogenic monogenic BRCA1/2 variant; for these individuals, the single pathogenic variant is the most important factor in determining their breast cancer risk. For the large majority of individuals who do not carry a pathogenic BRCA variant, the risk of breast cancer can be modeled by building a genetic risk score (GRS) which combines the small effects of many variants into a single score. Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease compared to the population baseline risk.

The breast cancer GRS was developed based on the summary statistics of a study that analyzed genomes of 228,951 individuals of primarily European ancestry, including 122,977 cases (individuals with breast cancer) and 105,974 healthy controls, provided by the Breast Cancer Association Consortium (BCAC).[9] The summary statistics from the meta-analysis were then adjusted for linkage disequilibrium using the LD-aware PRS-CS Bayesian shrinkage method.[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 predominantly European training data increases.[11] Orchid considers a GRS meaningfully predictive if individuals at approximately the 97.7th percentile have an odds ratio (OR) of at least 2. The breast cancer GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's breast cancer 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 females of British ancestry and defined breast cancer using the C50 ICD-10 code, yielding 14,001 cases and 206,854 controls (6.3% 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.

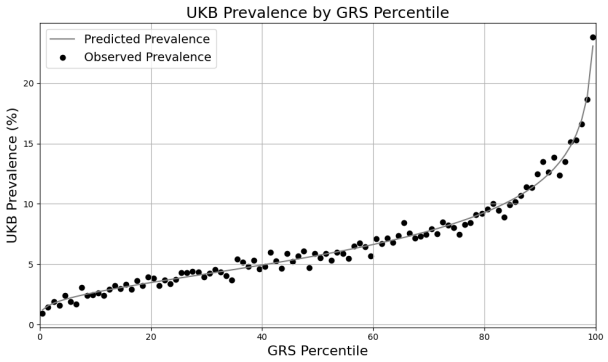

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

Table 1 shows the breast cancer 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 breast cancer, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	5.72%	1.00
Top 10%	15.08%	2.93
Top 5%	17.46%	3.49
Top 1%	23.92%	5.18

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

Estimating Lifetime Risk

The average observed prevalence of breast cancer in the UKB was 6.3%. This is considerably lower than the lifetime prevalence in the US general population, which has been estimated to be approximately 13%.[5] 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).

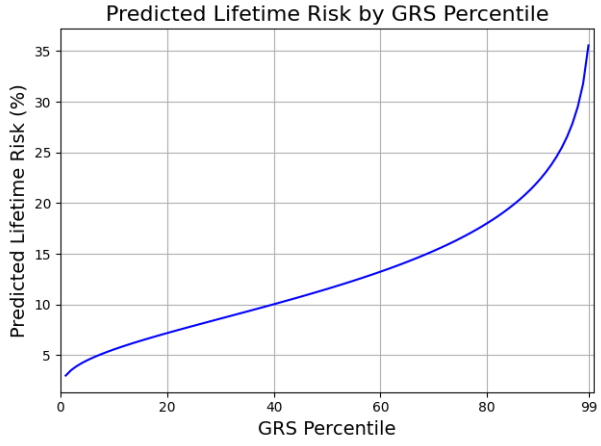

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 13% estimate from the National Cancer Institute.[5]

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	11.53%	1.00x
95th	26.55%	2.30x
97th	29.50%	2.56x
99th	35.56%	3.08x

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

Conclusion

In this study, we evaluated our breast cancer GRS, which was generated using summary statistics from BCAC, 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 breast cancer 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. Mayo Clinic Staff. Breast cancer – Symptoms and causes, July 26 2025. Accessed 2026. URL: https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470.
2. Centers for Disease Control and Prevention. Breast cancer risk factors, September 22 2025. U.S. Department of Health Human Services. URL: https://www.cdc.gov/breast-cancer/risk-factors/?CDC_AAref_Val=https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm.
3. J.D. Fackenthal and O.I. Olopade. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nature Reviews Cancer*, 7(12):937–948, 2007. doi:10.1038/nrc2054.
4. K.B. Kuchenbaecker, J.L. Hopper, D.R. Barnes, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*, 317(23):2402–2416, 2017. doi:10.1001/jama.2017.7112.
5. U.S. National Cancer Institute. Female breast cancer — Cancer Stat Facts, 2025. URL: https://seer.cancer.gov/statfacts/html/breast.html.
6. American Cancer Society. How common is breast cancer?, 2025. URL: https://www.cancer.org/cancer/types/

breast-cancer/about/how-common-is-breast-cancer.html.

7. American Cancer Society. Breast cancer survival rates by stage, 2025. URL: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

8. Mayo Clinic Staff. Breast cancer — Diagnosis and treatment, 2025. URL: https://www.mayoclinic.org/diseases-conditions/breast-cancer/diagnosis-treatment/drc-20352475.

9. K. Michailidou, S. Lindström, J. Dennis, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678):92–94, 2017. doi:10.1038/nature24284.

10. T. Ge, C.Y. Chen, Y. Ni, Y.A. Feng, and J.W. Smoller. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10:1776, 2019. doi:10.1038/s41467-019-09718-5.

11. F. 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. N. Chatterjee, J. Shi, M. 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.

Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
11.53%	1.86	3.46

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
6.3%	13.0%	8.65%

Table 4. Liability R^2. The estimated liability R^2 using a population prevalence of 13.0%.

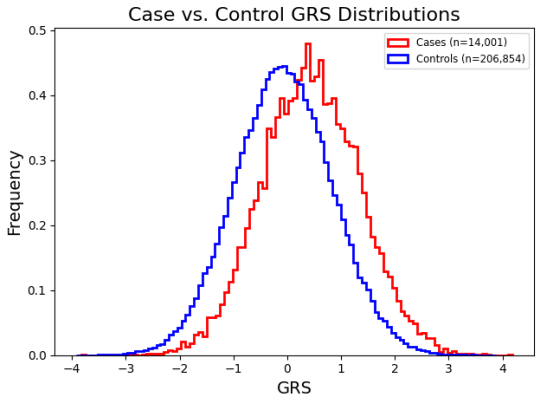

Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

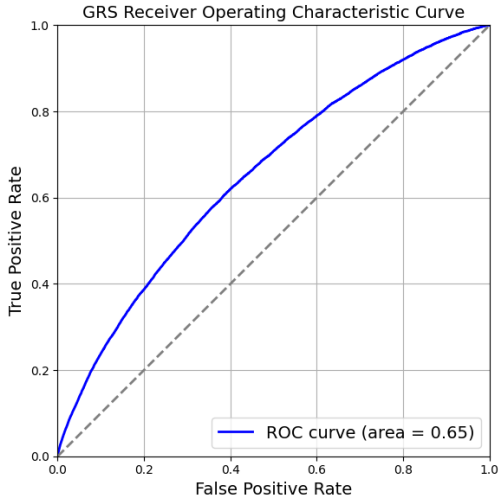

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds.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

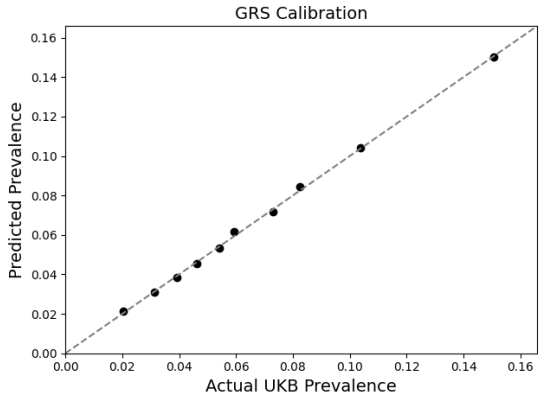

Figure 5. Calibration Curve. Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.