

Validating Orchid's Premature Menopause Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Menopause is the permanent cessation of menstruation caused by the loss of ovarian follicular activity. Most women experience natural menopause between the ages of 45 and 55 years,¹ but approximately 2% of women experience premature menopause,² which is defined as menopause before the age of 40. Premature menopause results in an earlier-than-expected loss of ovarian function and estrogen exposure, and can substantially shorten a woman's reproductive lifespan.

In addition to its effects on fertility, premature menopause is associated with increased risks of adverse long-term health outcomes, including osteoporosis and cardiovascular disease, and has also been associated with increased overall mortality.^{3,4} Hormone therapy is commonly recommended for women with premature loss of ovarian function, when appropriate, both to manage symptoms of estrogen deficiency and to reduce risks to bone and cardiovascular health.³

Genetic Risk Score

Premature menopause is shaped by both environmental and genetic factors. Although variants in individual genes have been associated with premature menopause, its inherited risk also has a polygenic component.⁵ Genetic risk scores (GRS), which combine the small effects of many variants into a single score, can be used to estimate genetic susceptibility. Although not diagnostic, a GRS can indicate how likely an individual is to experience premature menopause compared to the population baseline risk.

The premature menopause GRS was developed using data from Day et al. 2015⁶ and the UK Biobank (UKB).⁷ For the UKB, a subset of individuals was held out to use in the validation study. Summary statistics were generated for the UKB using the Regenie method.⁸ The Day et al. 2015 summary statistics were first imputed to the full variant panel using the summary-statistic imputation in SBayesRC,⁹ and summary statistics for both studies were then meta-analyzed using METAL.¹⁰ The final GRS was trained using SBayesRC and includes over 7 million variants.⁹

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 premature menopause GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's premature menopause GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.⁷ Because some data from the UKB was used to train the GRS, our validation study included only the set of samples that were held out for testing and not included in the GRS training. We restricted the analysis to females of British ancestry, and premature menopause was defined as follows:

- Field 2724 (had menopause): only women answering "Yes" were included. Women without menopause were excluded rather than treated as controls.
- Field 3591 (ever had hysterectomy): excluded if hysterectomy reported or status missing.
- Field 3581 (age at menopause): if reported at more than one visit, responses averaged when within 5 years of each other, otherwise excluded.
- Age at natural menopause restricted to 33–65 years (consistent with Ojavee et al. 2023¹²)
- Cases had an age at natural menopause under 40 years and controls 40 years or older.

This yielded 537 cases and 34,380 controls (1.5% prevalence). We then grouped individuals by GRS percentile and compared the observed premature menopause prevalence within each group to our model's predictions (Figure 1). For additional technical details, see the Supplementary Information.

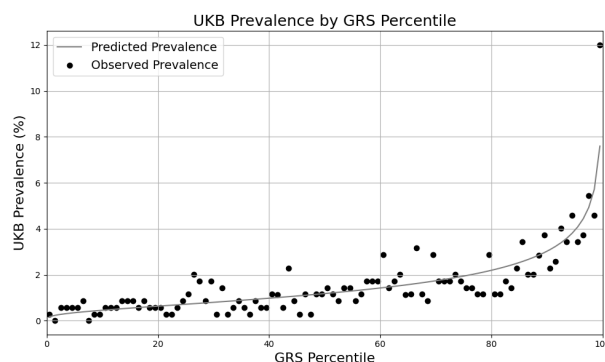


Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

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

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	1.29%	1.00
Top 10%	4.61%	3.70
Top 5%	5.84%	4.75
Top 1%	12.00%	10.44

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

Estimating Lifetime Risk

Mishra et al. estimate a 2% prevalence of premature menopause for women,² slightly higher than the observed prevalence of 1.5% in the UKB testing set. We adjust our model so that its average prevalence aligns with the Mishra et al. estimate (Figure 2).¹³ People at the high end of the GRS distribution are predicted to have an elevated risk of premature menopause relative to the population (Table 2).

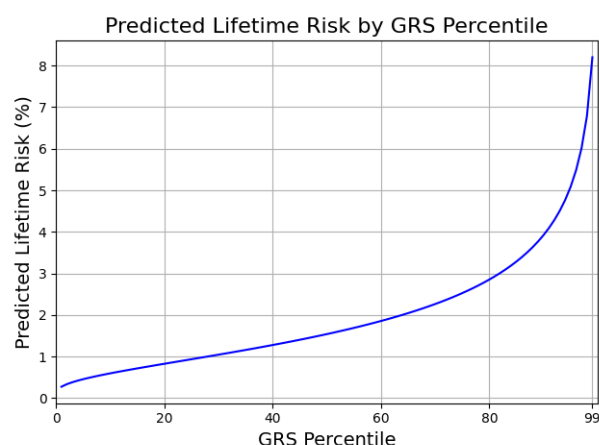


Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 2% estimate of Mishra et al.²

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	1.54%	1.00x
95th	5.09%	3.30x
97th	6.01%	3.90x
99th	8.20%	5.32x

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

Conclusion

In this study, we evaluated our premature menopause GRS on data from the UKB. We found that it performed well, particularly for identifying individuals with elevated risk of premature menopause relative to the population. In our embryo and couple reports, we adjust the model to predict risk consistent with the prevalence estimated by Mishra et al.² The premature menopause GRS model is available to individuals of all ancestry groups.

Acknowledgments

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

References

1. World Health Organization. Menopause. WHO Fact Sheet, 2024. URL: <https://www.who.int/news-room/fact-sheets/detail/menopause>.
2. G.D. Mishra, N. Pandeya, A.J. Dobson, et al. Early menarche, nulliparity and the risk for premature and early natural menopause. *Human Reproduction*, 32(3):679–686, 2017. doi:10.1093/humrep/dew350.
3. American College of Obstetricians and Gynecologists. Hormone therapy in primary ovarian insufficiency. Committee Opinion No. 698, 2017. URL: <https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2017/05/hormone-therapy-in-primary-ovarian-insufficiency>.
4. L.T. Shuster, D.J. Rhodes, B.S. Gostout, et al. Premature menopause or early menopause: long-term health consequences. *Maturitas*, 65(2):161–166, 2010. doi:10.1016/j.maturitas.2009.08.003.
5. K.S. Ruth, F.R. Day, J. Hussain, et al. Genetic insights into biological mechanisms governing human ovarian ageing. *Nature*, 596:393–397, 2021. doi:10.1038/s41586-021-03779-7.

6. F.R. Day, K.S. Ruth, D.J. Thompson, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics*, 47:1294–1303, 2015. doi:10.1038/ng.3412.
7. 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.
8. J. Mbatchou, L. Barnard, J. Backman, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53:1097–1103, 2021. doi:10.1038/s41588-021-00870-7.
9. 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.
10. C.J. Willer, Y. Li, and G.R. Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, 2010. doi:10.1093/bioinformatics/btq340.
11. F. Privé, H. Aschard, S. Carmi, 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. S.E. Ojavee, L. Darrous, M. Patxot, et al. Genetic insights into the age-specific biological mechanisms governing human ovarian aging. *American Journal of Human Genetics*, 110(9):1549–1563, 2023. doi:10.1016/j.ajhg.2023.07.006.
13. N. Chatterjee, J. Shi, and M. 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
1.54%	2.11	4.47

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
1.5%	2%	8.78%

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

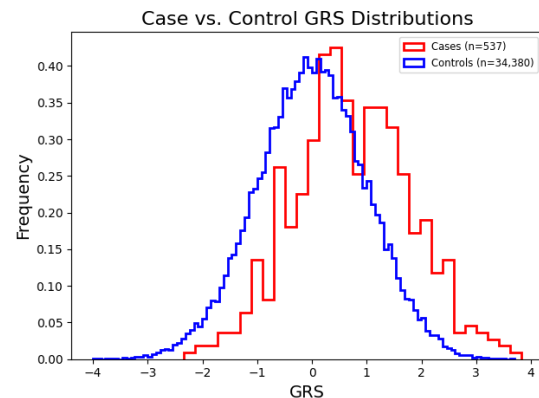


Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

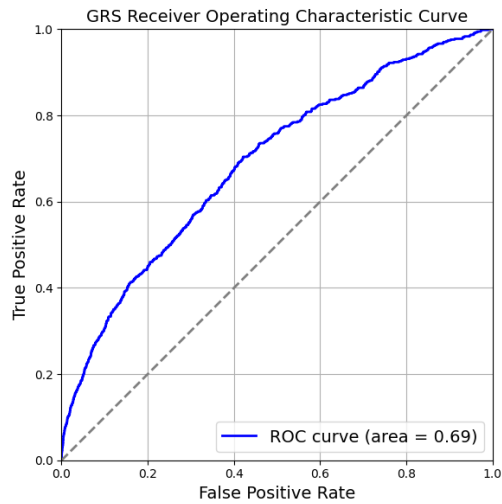


Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds.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

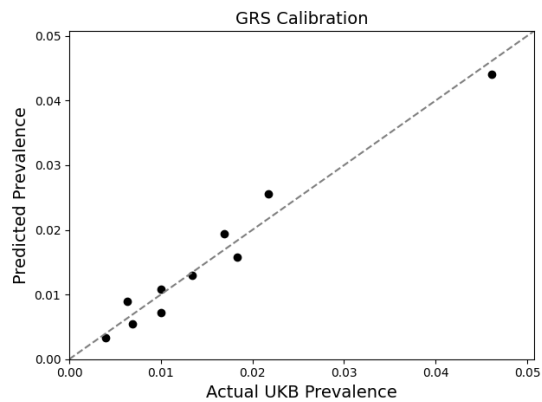


Figure 5. Calibration Curve. Calibration plot showing observed premature menopause prevalence versus predicted risk across GRS deciles.