

# Validating Orchid's Endometriosis Genetic Risk Score

Written by Orchid Bioinformatics Staff

## Introduction

Endometriosis is a chronic inflammatory disease characterized by the presence of uterine-like tissue outside the uterus, primarily within the pelvic region. It commonly causes severe pelvic pain, painful intercourse, and infertility. Symptoms can vary widely and are often accompanied by chronic fatigue. The disease usually occurs between menarche and menopause but can occasionally appear outside this range. Risk factors include early menarche, shorter menstrual cycles, low BMI, and family history.<sup>1</sup>

Endometriosis affects up to 10% of reproductive-age women worldwide, roughly 176 million individuals.<sup>1,2</sup> Most cases go undiagnosed because definitive diagnosis requires surgical confirmation, complicating efforts to obtain precise prevalence estimates. However, the condition is notably common among women experiencing infertility, affecting up to 50% in some populations studied.<sup>1</sup> Treatment of severe cases includes surgical removal of lesions and hormonal therapies to manage symptoms. Unfortunately, current therapies often have adverse effects, are contraceptive, or fail to provide lasting relief.<sup>1</sup>

## Genetic Risk Score

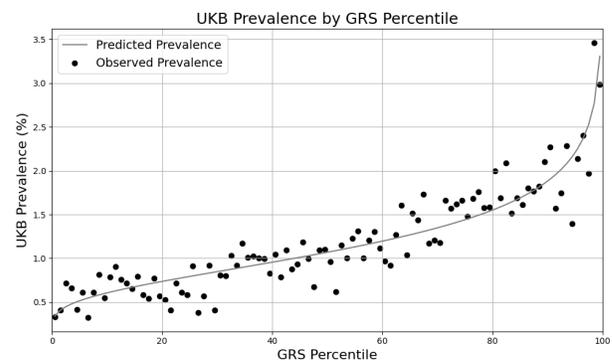
Endometriosis is shaped by both environmental and genetic factors. Monogenic testing is not available because no single gene causes the condition. Genetic risk scores (GRS), which combine the small effects of many variants into a single score, are currently the only way to estimate genetic risk. Although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

Orchid's endometriosis GRS was trained following current industry standards.<sup>3,4</sup> The GRS was constructed using the SBayesRC algorithm trained on publicly available FinnGen summary statistics.<sup>5,6</sup> The summary statistics include 20,190 cases and 130,160 controls.<sup>7</sup> The resulting GRS contains over a 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.<sup>8</sup> Orchid considers a GRS meaningfully predictive if individuals at roughly the 97.7th percentile have an odds ratio (OR) of at least 2. The endometriosis GRS meets this criterion for the European and Central South Asian ancestry groups and is available to individuals in these groups. Availability for an individual may vary due to admixture.

## Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's endometriosis GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.<sup>9</sup> We restricted the analysis to females of British ancestry and defined endometriosis as any diagnoses under ICD-10 codes N80.1 through N80.9, yielding 2,577 cases and 218,278 controls (1.2% 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 endometriosis 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 endometriosis, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

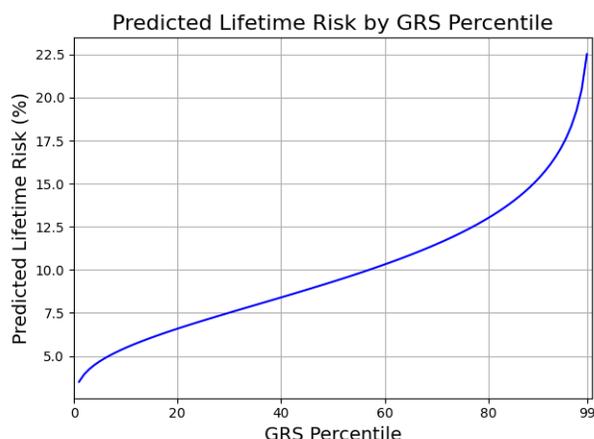
GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	1.01%	1.00
Top 10%	2.17%	2.18
Top 5%	2.55%	2.56
Top 1%	2.99%	3.02

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

## Estimating Lifetime Risk

The average observed prevalence of endometriosis in the UKB was 1.2%. This is considerably lower than the lifetime prevalence, which has been estimated to be approximately 10%.<sup>2</sup> 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.<sup>10</sup> 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).<sup>11</sup> 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 10% estimate.<sup>2</sup>

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	9.32%	1.00x
95th	17.65%	1.89x
97th	19.23%	2.06x
99th	22.52%	2.42x

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

## Conclusion

In this study, we evaluated our endometriosis 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 endometriosis GRS model is available to individuals of European and Central South Asian ancestry.

## Acknowledgments

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

## References

1. K. T. Zondervan, C. M. Becker, K. Koga, et al. Endometriosis. *Nat Rev Dis Primers*, 4:9, 2018. doi: 10.1038/s41572-018-0008-5.
2. B. Eskenazi and M. L. Warner. Epidemiology of endometriosis. *Obstet Gynecol Clin North Am*, 24(2):235–258, 1997. doi:10.1016/S0889-8545(05)70302-8.
3. 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.
4. 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.
5. 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. *Nat Genet*, 56:767–777, 2024. doi:10.1038/s41588-024-01704-y.
6. FinnGen. FinnGen Release 12 Summary Statistics. <https://console.cloud.google.com/storage/browser/finngen-public-data-r12>, 2025. Accessed December 5, 2025.
7. FinnGen. FinnGen Release 12 Phenotypes. <https://r12.finngen.fi/>, 2025. Accessed December 15, 2025.
8. 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.
9. 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.

10. 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. *Am J Epidemiol*, 186:1026–1034, 2017. doi: 10.1093/aje/kwx246.
11. N. Chatterjee, J. Shi, M. García-Closas, et al. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.

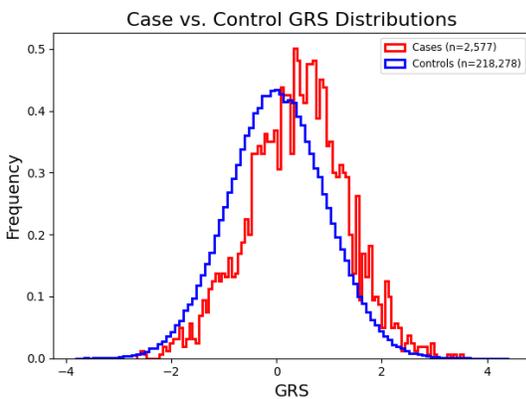
### Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
9.32%	1.56	2.45

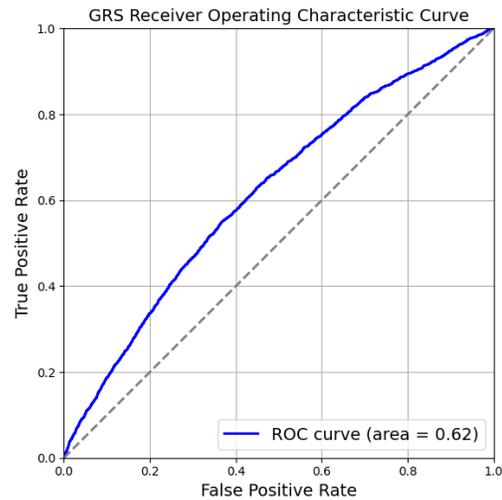
**Table 3. OR per SD.** The baseline risk for an individual with a median GRS, and the predicted OR at one and two SDs, respectively. A GRS must have a predicted OR >2 at 2 SD to be included in Orchid’s clinical reports.

UKB Prevalence	Population Prevalence	Liability R <sup>2</sup>
1.2%	10.0%	4.48%

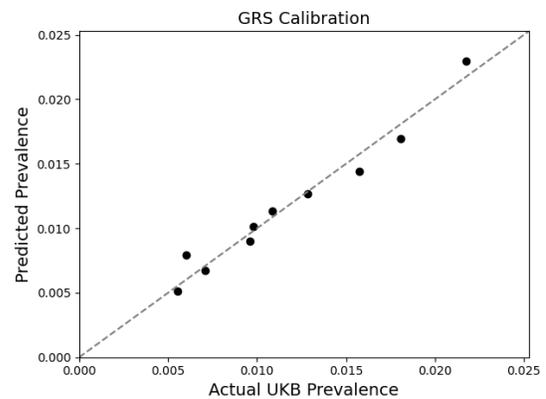
**Table 4. Liability R<sup>2</sup>.** The estimated liability R<sup>2</sup> using a population prevalence of 10%.



**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.