

# From Polygenic Scores to Phenotypic Screening: A Multi-Trait Framework for Cost-Free Risk Stratification in Endometriosis

Tarantino Barbara<sup>(1)</sup>, Gentilini Davide<sup>(1,2)</sup>

(1) Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

(2) Bioinformatics and Statistical Genomic Unit, IRCCS Istituto Auxologico Italiano, Milan, Italy

CORRESPONDING AUTHOR: Tarantino Barbara, [barbara.tarantino@unipv.it](mailto:barbara.tarantino@unipv.it)

## ABSTRACT

**Introduction:** Endometriosis is a chronic inflammatory condition affecting approximately 10% of women of reproductive age and is often diagnosed late due to nonspecific symptoms, overlap with common conditions such as primary dysmenorrhea, and the reliance on invasive laparoscopy [1,2]. Early detection could reduce patient burden and long-term complications, but current diagnostic tools remain limited. Genome-wide association studies (GWAS) have identified several genetic risk variants [3], yet their individual effects are modest. Polygenic risk scores (PRSs), which aggregate the effects of multiple variants, show promise but still lack the accuracy required for clinical application due to limited replication, small effect sizes, and population-specific variability [4,5]. Recent findings suggest that endometriosis is linked to a range of genetically influenced traits—such as immune, metabolic, and psychiatric characteristics—pointing toward the potential of multi-trait approaches to improve early, non-invasive risk stratification.

## OBJECTIVES

This study aims to develop a non-invasive, cost-effective strategy for endometriosis risk stratification using a genetics-informed, two-phase approach. First, we evaluated whether polygenic scores (PRSs) related to a broad spectrum of complex traits could predict disease risk and reveal genetically defined subgroups among patients. Next, we identified the most informative traits associated with these genetic risk profiles and translated them into a targeted phenotypic questionnaire. We then assessed whether this phenotype-only model could accurately classify endometriosis cases, offering a feasible alternative to genetic testing for early detection in real-world settings.

## METHODS

We analyzed 1,996 genotyped women (862 cases, 1,134 controls) and computed 4,490 PRSs across complex traits. After filtering and trait mapping, 645 scores were retained; one per trait was selected via bootstrap logistic regression (218), then reduced to 40 via LASSO. Supervised machine learning models (logistic regression, random forest, XGBoost, neural networks) [6,7] were trained to evaluate the predictive performance of the PRS-based model. Top-ranking PRSs from the best-performing model were used to cluster endometriosis cases, identifying genetically defined subgroups. Traits linked to these PRSs were used to design a targeted phenotypic questionnaire. The questionnaire was tested in an independent cohort ( $n = 506$ ), where curated phenotypic features were used to train classification models. The best model was then used to generate a non-invasive, phenotype-only risk score for endometriosis stratification.

## RESULTS

The multi-PRS model significantly outperformed the endometriosis-specific PRS (AUC = 0.636 vs. 0.546,  $p < 0.001$ ), with key contributions from traits related to height, early menarche, schizophrenia, and autoimmune disorders. Clustering based on the most informative PRSs identified two genetically defined subgroups with distinct clinical characteristics, including differences in endometrioma prevalence, gastrointestinal symptoms, and disease stage. A phenotype-only model trained on questionnaire data demonstrated high discriminative ability (AUC = 0.904), with CA125, fatigue, gynecological symptoms, and muscle pain emerging as the most informative features, supporting its potential as a cost-effective and non-invasive tool for early risk stratification.

## CONCLUSIONS

Our results demonstrate that leveraging polygenic information to identify trait-level predictors enables the development of accurate, phenotype-based models for endometriosis risk stratification. The use of AI-driven approaches allows robust prediction from a minimal set of non-invasive, low-cost clinical features—reducing reliance on genetic testing and supporting more accessible, early diagnostic strategies within precision gynecology.

## BIBLIOGRAPHY:

1. Zondervan K.T., Becker C.M., Missmer S.A., Endometriosis. *N Engl J Med.*, 2020 Mar; 382(13):1244–56
2. Vannuccini S., Lazzeri L., Orlandini C. et al., Mental health, pain symptoms and systemic comorbidities in women with endometriosis: a cross-sectional study. *J Psychosom Obstet Gynaecol.*, 2018 Dec; 39(4):315–20
3. Sapkota Y., Steinhorsdottir V., Morris A.P. et al., Meta-analysis identifies five novel loci associated with endometriosis. *Nat Commun.*, 2017 May; 8:15539
4. Kløve-Mogensen K., Rohde P.D., Twisstmann S. et al., Polygenic Risk Score Prediction for Endometriosis. *Front Reprod Health.*, 2021; 3:793226
5. Chatterjee N., Shi J., García-Closas M., Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet.*, 2016 Jul; 17(7):392–406
6. Bendifallah S., Puchar A., Suisse S. et al., Machine learning algorithms as new screening approach for patients with endometriosis. *Sci Rep.*, 2022 Jan; 12(1):639
7. Miotto R., Wang F., Wang S., Jiang X., Dudley J.T., Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.*, 2018 Nov; 19(6):1236–46