

Nordic AI Meet 2026

P2: AI in Health, Medicine, Care & Human Well-being

7. Data-Driven Multidimensional Risk Stratification of PCOS/PMOS Using Population-Scale Health Data

Authors: Priyanka Mehra¹; Anagha Joshi¹

Affiliations: 1. Faculty of Medicine, Department of Clinical Science, University of Bergen, Norway

Abstract: Polycystic Ovary Syndrome (PCOS), recently reframed as Polycystic Metabolic Ovary Syndrome (PMOS) to emphasize its metabolic component, is a heterogeneous condition involving ovarian, endocrine, and metabolic dysfunction. This heterogeneity poses a challenge for population-scale risk assessment, particularly when complete clinical diagnostic information is unavailable. We developed an interpretable, data-driven PMOS/PCOS-like risk stratification framework using nationally representative health data from women in India.

Ovarian, endocrine, and metabolic domain scores were constructed from clinically relevant indicators available in the National Family Health Survey (NFHS-5) and integrated into a continuous composite score. Women were subsequently stratified into low-, moderate-, and high-risk groups. We evaluated whether this representation captured meaningful differences in health burden by examining associations with noncommunicable diseases (NCDs), primary infertility, reproductive and maternal outcomes, blood pressure, hemoglobin, intimate partner violence (IPV), and sociodemographic and geographic characteristics using weighted population-level analyses and multivariable regression.

The framework classified 65% of women as low risk, 31% as moderate risk, and 4% as high risk. Higher PMOS/PCOS-like risk was associated with multiple adverse health outcomes. Compared with low-risk women, high-risk women had 33% higher adjusted odds of miscarriage, 70% higher odds of stillbirth, and 73% higher odds of caesarean delivery. Higher risk was also associated with elevated blood pressure and greater odds of several NCDs, particularly diabetes and thyroid disorder. High-risk women showed distinct socioeconomic and geographic patterning, with higher prevalence among urban, more educated, and wealthier women and in states with higher Human Development Index. Differences in IPV were also observed across risk groups.

These findings demonstrate how heterogeneous reproductive, endocrine, and metabolic information from population-scale health data can be integrated into an interpretable representation for women's health risk stratification. The framework provides a foundation for future machine-learning approaches to learn domain contributions and interactions and to develop and validate individualized PMOS/PCOS risk prediction using clinically characterized and longitudinal data.