



Research Article

Open Access

Polygenic Risk Scores (PRS/PGS) across Multi-ancestry and Cross-domain Settings: Statistical Framework, Methodological Advances, and Robustness Evaluation

Xuanjun Fang 

Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025

 Corresponding author: xuanjunfang@hitar.orgGenomics and Applied Biology, 2026, Vol.17, No.3 doi: [10.5376/gab.2026.17.0012](https://doi.org/10.5376/gab.2026.17.0012)

Received: 06 Apr., 2026

Accepted: 12 May, 2026

Published: 25 May, 2026

Copyright © 2026 Fang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Fang X.J., 2026, Polygenic risk scores (PRS/PGS) across multi-ancestry and cross-domain settings: statistical framework, methodological advances, and robustness evaluation, Genomics and Applied Biology, 17(3): 138-153 (doi: [10.5376/gab.2026.17.0012](https://doi.org/10.5376/gab.2026.17.0012))

Abstract Polygenic risk scores (PRS/PGS) aggregate genome-wide association study (GWAS) effect sizes to quantify individual-level genetic susceptibility, serving as a key bridge between genetic association findings and practical applications. With the rapid expansion of large-scale genotype-phenotype datasets, PRS methodology has evolved from early clumping-and-thresholding (C+T) approaches to frameworks that explicitly model linkage disequilibrium (LD) and effect size distributions using Bayesian shrinkage and penalized regression, and further incorporate functional annotations, multi-ancestry data, and transfer learning to improve predictive performance and interpretability. However, the portability and robustness of PRS across populations remain major challenges, often manifesting as reduced predictive accuracy, calibration bias, and unstable decision thresholds. From a statistical perspective, these issues can be understood as an estimand mismatch arising from differences in LD structure, allele frequency spectra, and effect distributions across populations. In this study, we revisit PRS within a unified statistical genetics framework by conceptualizing it as a model-dependent predictive functional, and link it to SNP heritability as part of a continuous inference chain from variance decomposition to individual-level prediction. Building on this perspective, we systematically review and compare state-of-the-art PRS methods under multi-population settings, including LD-aware Bayesian shrinkage, functionally informed models, multi-ancestry transfer learning, and model stacking and recalibration strategies, with representative methods such as PRS-CSx, CT-SLEB, and PolyPred. We further propose a standardized analytical workflow of “training-validation-freezing-external evaluation” and advocate a multi-dimensional evaluation framework based on relative R^2 /AUC, calibration metrics, and decision-curve net benefit. In addition, we discuss joint modeling of PRS with environmental and lifestyle factors and its applications in both human health and crop breeding. Finally, we address issues of cross-population inequity and ethical governance, and propose an integrated framework centered on multi-ancestry data expansion, causal and functional annotation integration, ancestry-aware modeling, environment coupling, and population-specific recalibration. This framework aims to promote PRS/PGS from a predictive tool toward a transferable, interpretable, and equitable decision-support system, providing a systematic foundation for the application of complex trait genetics.

Keywords Polygenic risk scores; Statistical genetics; SNP heritability; Cross-population prediction; Linkage disequilibrium; Bayesian shrinkage; Transfer learning; Gene-environment interaction; Fairness

Polygenic risk scores (PRS/PGS) aggregate effect size estimates derived from genome-wide association studies (GWAS) to generate individual-level predictions, thereby transforming locus-trait associations into quantitative measures of genetic susceptibility for complex traits or diseases. With the continuous expansion of large-scale genotype-phenotype cohorts and advances in computational methods, PRS construction has evolved from early clumping-and-thresholding (C+T) approaches to frameworks that explicitly model linkage disequilibrium (LD) and effect size sparsity using Bayesian shrinkage and penalized regression. More recently, these methods have further incorporated functional annotations, fine-mapping, and multi-trait information to enhance signal-to-noise ratio, interpretability, and predictive performance. This methodological progression reflects a paradigm shift in statistical genetics from hypothesis-driven analyses to genome-wide, hypothesis-free scanning (Cai et al., 2021; Weissbrod et al., 2022; Zhang et al., 2023; Fang and Wu, 2026). Meanwhile, multi-ancestry training and cross-population transfer learning approaches have rapidly developed, and the increasing availability of