From a statistical genetics perspective, PRS is not merely a predictive tool but a model-dependent predictive function, whose performance and scope are jointly determined by training data, LD structure, and effect estimation methods. In correspondence with SNP heritability as a measure of variance explained, PRS represents the expression of genetic signal at the individual prediction level. Together, they form a variance-prediction duality. Within this framework, the decline in cross-population predictive performance can be understood as an estimand mismatch arising from differences in LD structure, allele frequency spectra, and effect distributions across populations.

At present, cross-population robustness remains the central bottleneck for translating PRS/PGS into clinical and agricultural practice. Advances in multi-ancestry data resources and methodology are progressively addressing this challenge. Multi-ancestry training, hierarchical modeling, ancestry-aware LD modeling, and the integration of local ancestry information enable improved balance between shared and population-specific effects. The incorporation of functional annotation and causal refinement further reduces the impact of tagging effects on cross-population prediction. At the application level, adherence to a standardized workflow-"training-validation-freezing-external evaluation" -together with multi-dimensional performance metrics for predictive accuracy and fairness, is increasingly recognized as best practice.

Looking forward, the development of PRS/PGS is expected to follow three major directions. First, integration of causal inference and functional annotation will enhance signal identification and cross-population robustness through structured priors. Second, multi-ancestry modeling and transfer learning will reduce performance gaps in underrepresented populations. Third, coupling with environmental and lifestyle factors will extend predictive functions into context-dependent models through explicit modeling of gene-environment interaction. In medicine, these advances will support stratified screening and personalized interventions for high-burden diseases; in crop breeding, they will enable efficient strategies combining general adaptability with environment-specific optimization.

At the same time, several key challenges remain. These include imbalances in multi-ancestry data and annotation resources, limited coverage of low-frequency and structural variants, cross-platform batch effects, and residual population structure, as well as ethical and governance considerations such as data sovereignty and fairness. With the development of pangenome reference systems, long-read sequencing, and multi-omics integration, along with the establishment of open benchmarks, standardized quality control, and transparent reporting frameworks, PRS/PGS are expected to evolve into a general predictive platform that is scientifically robust and socially responsible, with broad applications in global health and food security.

Author Contributions

Xuanjun Fang conducted the study, including literature review, data analysis, and drafting and revising the manuscript. The author has read and approved the final version of the manuscript.

Acknowledgements

This work was supported by the Major Program of the National Natural Science Foundation of China (Grant No. 30490254).

References

Adeyemo A., Balaconis M., Darnes D., and Zhou A., 2021, Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps, Nature Medicine, 27(11): 1876-1884.
https://doi.org/10.1038/s41591-021-01549-6

Alemu A., Åstrand J., Montesinos-López O., Sánchez J., Fernández-González J., Tadesse W., Vetukuri R., Carlsson A., Ceplitis A., Crossa J., Ortiz R., and Chawade A., 2024, Genomic selection in plant breeding: Key factors shaping two decades of progress, Molecular Plant, 17(4): 552-578.
https://doi.org/10.1016/j.molp.2024.03.007

Andreoli L., Peeters H., Van Steen K., and Dierickx K., 2024, Taking the risk: Ethical reasons and moral arguments in the clinical use of polygenic risk scores, American Journal of Medical Genetics Part A, 194(7): 1939-1956.
https://doi.org/10.1002/ajmg.a.63584