

4.3 Mitigation strategies: a multi-layer framework from data to governance

Addressing inequity in cross-population PRS/PGS applications requires systematic improvements at three levels: data, methodology, and governance.

At the data level, efforts should focus on expanding multi-ancestry and multi-ecological GWAS datasets and reference panels to improve coverage of low-frequency and structural variants, as well as developing comprehensive functional annotation resources across tissues and environments. Standardization of data quality control (QC), genomic coordinates, and allele coding conventions is essential to reduce technical heterogeneity across studies (Zhang et al., 2023; Kachuri et al., 2024; Kullo, 2024). In addition, promoting cross-institutional data sharing and establishing standardized consortia, along with dynamically updated public resource repositories, will improve both representativeness and accessibility of data.

At the methodological level, advances are needed in multi-ancestry joint modeling, hierarchical modeling, ancestry-aware LD modeling, and transfer learning approaches (e.g., PRS-CSx, CT-SLEB, Joint-Lassosum). Incorporating local ancestry information and domain adaptation techniques can further enhance model adaptability to population structure differences. Small-sample reweighting or recalibration in the target population has been shown to effectively reduce performance gaps in practice (Zhang et al., 2023; Kachuri et al., 2024). Furthermore, integrating functional annotations and causal inference approaches as biologically informed priors can reduce noise and improve robustness in cross-population prediction.

At the evaluation and governance level, it is essential to establish a systematic “fairness metric framework,” reporting performance metrics separately across populations, including AUC, R^2 , calibration slope and intercept, Brier score, net benefit, and reclassification metrics (NRI). Performance differences should be quantified using measures such as the transferability ratio (T) and Δ AUC. In addition, decision thresholds and rules should be optimized for each population, and uncertainty measures (e.g., confidence intervals and calibration curves) should be reported to avoid overinterpretation of single metrics. Adoption of transparent reporting standards (e.g., Polygenic Score Catalog, PRS Reporting Statement) and regulatory and ethical compliance frameworks (Wand et al., 2020; Lewis et al., 2024; Xiang et al., 2024) will further ensure that PRS/PGS applications are scientifically robust, interpretable, and equitable.

5 Discussion

From a unified statistical inference framework, the performance differences among existing PRS/PGS methods fundamentally arise from distinct modeling assumptions regarding effect size distributions, linkage disequilibrium (LD) structure, and sparsity, thereby corresponding to different statistical targets (estimands). From this perspective, methodological evolution can be viewed as a progressive approximation to the question of how GWAS-derived effects can be transformed into stable predictive functions.

Compared with baseline approaches such as clumping and thresholding (C+T), LD-aware Bayesian shrinkage methods (e.g., LDpred2 and PRS-CS) apply continuous shrinkage to effect sizes under LD constraints, typically achieving higher predictive performance and smoother behavior with respect to hyperparameters under the same training data and reference panels. When functional annotations are incorporated (e.g., LDpred-funct and AnnoPred), differential shrinkage or preferential inclusion of variants in functionally enriched regions can further improve the signal-to-noise ratio and partially mitigate cross-population bias induced by tagging effects (Sima et al., 2024). In settings involving heterogeneous data or extrapolation, multi-ancestry and cross-population approaches (e.g., PRS-CSx, hierarchical modeling, and model stacking) balance shared and population-specific effects, often achieving a better trade-off between transferability and stability (Zhang et al., 2023).

Based on these insights, a practical three-step decision framework can be adopted, centered on “scenario-resource-validation.” First, modeling strategies should be selected based on the sample size and representativeness of the target population (population-specific models versus multi-ancestry transfer models). Second, method selection should consider the availability of LD reference panels and functional annotations