

errors introduced by sample selection bias may be significantly amplified. Therefore, in interpreting SNP heritability estimates, it is important to distinguish between its relative stability as a baseline parameter and its potential propagation effects in downstream analyses, in order to avoid overgeneralization of research conclusions.

3.6 Summary of results

Through a comparative analysis integrating multiple methods and data sources, several consistent conclusions can be drawn. First, systematic differences exist among estimation methods, with biases generally ranging from −10% to +40%, indicating that method selection itself constitutes a major source of variation in results. Second, individual-level approaches represented by GREML tend to provide relatively higher and more stable heritability estimates, whereas summary-statistics-based methods such as LDSC commonly exhibit a tendency toward underestimation. In contrast, SumHer, by explicitly modeling linkage disequilibrium (LD) structure and allele frequency distributions, can improve the plausibility of estimates to some extent.

More importantly, these differences are not incidental but reflect the dependence of SNP heritability on multiple structural factors. These factors primarily include the complexity of LD structure, the genomic coverage of SNP markers, and the fundamental assumptions of the models employed. Therefore, SNP heritability should not be interpreted as a single fixed value, but rather understood within the context of specific data structures and analytical frameworks.

4 Discussion

4.1 Estimand mismatch as the fundamental source of discrepancy

The central finding of this study can be summarized as a methodological principle: SNP-based heritability estimates obtained from different methods do not correspond to the same statistical quantity, but rather to distinct estimands defined by data structure and model assumptions (estimand mismatch). This interpretation is consistent with recent statistical frameworks of SNP heritability, which emphasize that different models target different estimands rather than a single underlying biological parameter (Fang, 2026; Fang and Wu, 2026).

This perspective provides a unified explanation for the systematic differences observed in large-scale datasets such as the UK Biobank. Specifically, GREML-based methods typically yield higher estimates, LDSC-based approaches tend to underestimate heritability, and SumHer can produce substantially higher estimates under certain conditions (Hou et al., 2019; Speed et al., 2020).

From a statistical standpoint, SNP-based heritability is not a fixed “true parameter,” but a conditional quantity that can be expressed as:

$$h_{\text{SNP}}^2 = \frac{\text{Var}(\hat{g}|\text{model, SNP set, LD})}{\text{Var}(y)}$$

Accordingly, differences across methods do not represent contradictions, but rather reflect alternative modeling perspectives on genetic variance (Rawlik et al., 2020).

4.2 Statistical origins of method-dependent differences

Differences among heritability estimation methods primarily arise from variations in data representation and the efficiency with which information is utilized. GREML relies on individual-level genotype data, enabling the direct construction of a genetic relationship matrix (GRM) among individuals and the estimation of genetic effects within a variance component framework; consequently, it makes more comprehensive use of available information. In contrast, LDSC and SumHer are based mainly on GWAS summary statistics. Their analytical objects are no longer the complete genotype structures of individuals but rather statistical results compressed through marginal association analyses. Although such approaches offer clear advantages for integrating large-scale public datasets, this compression inevitably weakens certain covariance structures present at the individual level, potentially leading to reduced estimation efficiency and increased bias. Previous studies have shown that, under identical data conditions, summary-based methods generally exhibit higher variance and greater susceptibility to bias compared with individual-level methods (Bulik-Sullivan et al., 2015; Ni et al., 2018).