

Therefore, the systematic underestimation of heritability by LDSC is not merely attributable to computational error but is closely related to information loss inherent in its data input format.

Secondly, methodological differences are also associated with the assumptions each model makes about the distribution of genetic effects. GREML typically assumes that all SNPs contribute equally to the variance, while LDSC further simplifies the relationship between genetic effects and LD scores into a linear structure. In contrast, the LDAK framework underlying SumHer allows SNP effects to vary with allele frequency and LD structure. The key issue is that real genetic architectures often deviate from the assumption of homogeneous effects: low-frequency variants may have larger effects, and regions with low LD may harbor a higher concentration of causal variants. Under such circumstances, both standard GREML and LDSC may underestimate heritability, whereas LDAK, by incorporating MAF- and LD-based weighting, can to some extent improve the model's fit to the true genetic architecture (Speed et al., 2017; Speed and Balding, 2019). Thus, differences in results across methods should be understood as reflecting differences in how well each model captures the underlying genetic architecture, rather than as mere random estimation error.

The structure of linkage disequilibrium and its regional heterogeneity further amplify these methodological differences. Analyses based on the UK Biobank have shown that genetic variance is not uniformly distributed across the genome but may be concentrated in specific high-LD regions. For example, the MHC region exhibits extremely strong LD, and when this region is excluded, the estimated SNP heritability for certain traits can decrease by more than 0.2 (Ge et al., 2017). This finding indicates that SNP heritability is not a direct measure of total genetic variance but rather reflects the portion of genetic variance that can be captured by observed SNPs under specific marker density and LD coverage conditions. In other words, SNP heritability inherently has a pronounced “LD-weighted” property, with its magnitude depending on whether causal variants are effectively tagged by existing markers, rather than solely on the intrinsic genetic basis of the trait.

Based on the above analysis, this study emphasizes the concept of “capturability.” SNP heritability is not equivalent to true narrow-sense heritability and should not be simply interpreted as:

$$h_{\text{SNP}}^2 \neq h_{\text{true}}^2$$

More precisely, it represents the genetic variance explained by observed SNPs through linkage disequilibrium (LD) tagging:

$$h_{\text{SNP}}^2 = \text{variance explained by SNPs through LD tagging}$$

This understanding is consistent with previous studies, which indicate that SNP heritability reflects only the genetic variation that is “tagged” by the observed markers (Yang et al., 2015). From this perspective, so-called “missing heritability” does not necessarily imply that genetic effects are truly absent, but is more likely the result of insufficient SNP coverage, incomplete LD tagging, and limitations imposed by model assumptions acting in combination.

4.3 Implications for comparability and interpretation

Focusing on the issue of estimand mismatch, a key conclusion can be further clarified: SNP heritability estimates obtained from different methods are, in most cases, not strictly statistically comparable. The notion of “comparison” theoretically presupposes that the estimands targeted by different methods are identical; however, this assumption is often difficult to satisfy in practical applications. Differences in the composition of SNP sets (such as variations in marker density), discrepancies in the sources of linkage disequilibrium (LD) reference panels (e.g., 1000 Genomes versus in-sample LD), and differing model assumptions regarding effect size distributions (such as uniform distribution assumptions versus models weighted by LD or minor allele frequency, MAF) all alter the definition of the estimand itself (Hou et al., 2019; Speed et al., 2020). When these conditions are not rigorously standardized, horizontal comparisons of estimated values lack statistical validity.

This perspective helps to reinterpret the frequently observed “inconsistencies” in the existing literature. Conventional explanations often interpret the lower estimates obtained from LDSC as evidence of “missing