

statistical quantities (estimands) they target, and these differences directly influence their applicability and interpretability across different research contexts.

2.2.1 GREML framework (GCTA)

The GREML (Genomic Restricted Maximum Likelihood) approach is based on a linear mixed model (LMM) that estimates genetic variance using genomic similarity between individuals (Yang et al., 2016). The statistical interpretation and estimand definition of GREML have been discussed in detail in previous work (Fang, 2026). The model is specified as:

$$y = X\beta + g + \varepsilon$$

where: y : phenotype vector; X : covariate matrix (including age, sex, and principal components)' β : fixed effects; g : genetic random effects; ε : residual environmental effects.

The random effects are assumed to follow:

$$g \sim N(0, \sigma_g^2 G), \varepsilon \sim N(0, \sigma_e^2 I)$$

where G is the genomic relationship matrix (GRM), constructed from genome-wide SNPs to capture genetic similarity between individuals (Yang et al., 2010).

SNP-based heritability is defined as:

$$h_{\text{SNP}}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

The corresponding estimand represents the proportion of additive genetic variance captured by observed SNPs through linkage disequilibrium.

Under large sample sizes and correct model specification, GREML provides asymptotically unbiased and efficient estimates (Hou et al., 2019). Extensions such as GREML-LDMS, which stratify SNPs by MAF and LD to construct multiple GRMs, can further improve estimation accuracy and mitigate model misspecification (Speed et al., 2017).

2.2.2 LD score regression (LDSC)

LD Score Regression (LDSC) estimates SNP-based heritability using GWAS summary statistics by exploiting the relationship between association test statistics and LD scores (Bulik-Sullivan et al., 2015).

The fundamental model is:

$$E[\chi_j^2] = 1 + \frac{N h^2 l_j}{M}$$

where: χ_j^2 : association test statistic for SNP j ; l_j : LD score (sum of squared correlations with neighboring SNPs); N : sample size; M : total number of SNPs.

The advantages of LDSC (Linkage Disequilibrium Score Regression) primarily lie in its dual optimization of data dependency and statistical inference capability. This method is based on GWAS summary statistics and does not require access to individual-level data. On this basis, LDSC can conveniently integrate results from different study cohorts, demonstrating strong adaptability within large-scale meta-analysis frameworks. More importantly, by modeling the structure of linkage disequilibrium, the method effectively distinguishes confounding effects due to population structure from genuine polygenic genetic signals.

However, LDSC relies on external LD reference panels (e.g., 1 000 Genomes), and mismatches between reference and target populations may introduce systematic bias (Ni et al., 2018). Moreover, LDSC implicitly assumes homogeneous SNP effect sizes, which is often violated in realistic genetic architectures, leading to underestimation of heritability (Speed et al., 2020).