

3.2 Quantitative comparison and relative differences

To quantitatively assess differences across methods, we used the GRE (closed-form estimator; $h^2 \approx 0.60$) as the reference baseline and computed relative deviations as:

$$\text{Relative Difference} = \frac{h_{\text{method}}^2 - h_{\text{GRE}}^2}{h_{\text{GRE}}^2}$$

From the results, different methods exhibit directional bias patterns. Among them, S-LDSC produces estimates that are generally lower than the baseline level, with a deviation of approximately -7% , indicating a certain degree of systematic underestimation. This phenomenon is typically associated with its simplified modeling of linkage disequilibrium (LD) structure and its treatment of pleiotropic signals. In contrast, the estimates obtained from SumHer are slightly higher than the GRE baseline, with a deviation of about $+5\%$. Although this does not represent a substantial departure, it still reflects a mild inflation effect arising from its model assumptions or weighting scheme. Furthermore, results from GREML-type methods show more pronounced variability, with deviations ranging from approximately $+14\%$ to $+37\%$. This fluctuation is clearly dependent on specific model settings and SNP coverage density, suggesting a high sensitivity to data structure and parameter configuration.

When these findings are considered in the context of existing large-scale comparative studies, their overall trends appear to be largely consistent. In large datasets such as the UK Biobank, LDSC-type methods generally exhibit underestimation in the range of approximately 7% to 14% , whereas SumHer may show varying degrees of overestimation within a range of about 5% to 38% . Meanwhile, due to differences in implementation strategies and modeling details, GREML-type methods tend to display a certain degree of variability in their estimates across different studies (Hou et al., 2019; Speed et al., 2020).

3.3 Key empirical observations

3.3.1 GREML-family methods produce higher estimates

From existing empirical evidence, GREML-type methods based on individual-level data (such as GCTA, GRE, and moment estimators) generally yield relatively higher estimates of SNP heritability. This pattern is not incidental, but is closely related to their methodological characteristics. First, these approaches directly utilize the full genotype matrix for modeling, thereby avoiding information loss that may occur during data compression or summarization. Second, by constructing a genetic relationship matrix (GRM), the model can explicitly incorporate linkage disequilibrium (LD) structure, allowing for a more comprehensive representation of correlations among loci. Third, in terms of statistical efficiency, the use of individual-level data enables more effective utilization of available information in parameter estimation. For these reasons, heritability estimates obtained from GREML-type methods are closer to the range of true genetic variance that can be captured by the current set of SNPs under the constraints of LD structure (Yang et al., 2010; Hou et al., 2019). As sample sizes increase beyond 100,000, GREML estimates exhibit markedly improved stability, accompanied by a substantial reduction in standard errors, indicating greater statistical reliability of the estimates (Ge et al., 2017).

3.3.2 LDSC systematically underestimates SNP heritability

In contrast to GREML-type methods, LD score regression (LDSC) and its extensions (e.g., S-LDSC), which are based on summary statistics, tend to yield systematically lower estimates of heritability in most studies, with the magnitude of bias typically ranging from approximately -7% to -14% . This underestimation can be explained from multiple perspectives.

First, LDSC relies on external LD reference panels (such as the 1000 Genomes Project), and discrepancies in genetic structure between the reference population and the target sample may lead to mismatches in LD estimation, thereby introducing systematic bias (Ni et al., 2018). Second, at the level of model assumptions, LDSC generally assumes that all SNPs contribute equally to genetic variance; however, a substantial body of evidence indicates that the true genetic architecture is often jointly influenced by minor allele frequency (MAF) and LD structure. This simplifying assumption therefore limits its ability to accurately capture the genetic basis of complex traits (Speed et al., 2020). In addition, because LDSC relies solely on summary statistics for inference, the covariance