

thereby exhibiting method-dependent statistical interpretations (Rawlik et al., 2020). The GREML approach, which constructs a genetic relationship matrix (GRM) based on individual-level data, yields estimates that reflect the genetic variance components within this matrix framework. In contrast, LDSC relies on linkage disequilibrium (LD) structure to weight genome-wide effects, and its estimates are more akin to an LD-weighted average effect variance. Building upon this, SumHer introduces joint weighting based on minor allele frequency (MAF) and LD structure, thereby recharacterizing genetic variance and allowing its estimates to capture differential contributions from variants of varying frequencies. It is precisely these systematic differences in weighting schemes and model assumptions that lead to non-negligible discrepancies in both the numerical values and interpretative meanings of heritability estimates across methods, forming the key starting point for the subsequent comparative analyses and theoretical discussions in this study.

3 Results: UK Biobank Case Study and Quantitative Comparison

3.1 SNP-based heritability estimates across methods

Using European-ancestry samples from the UK Biobank (UKB), we systematically compiled and compared SNP-based heritability estimates for height across different methods. These approaches differ substantially in sample size, data representation, and model assumptions. The results are summarized in Table 1.

Table 1 SNP heritability estimates for height in UK Biobank

Method	Sample size	SNPs	h^2_{SNP}	SE	Relative difference
GREML (GCTA)	~20 000	~50k	0.50	0.02	baseline
GREML-like (moment estimator)	152 736	genome-wide	0.685	0.004	+37%
GRE(closed-form estimator)	290 000	~460k	0.60	-	baseline
S-LDSC	same	same	0.56	-	-7%
SumHer (LDAK)	same	same	0.63	-	+5%

Note: Data compiled from UKB-based empirical studies

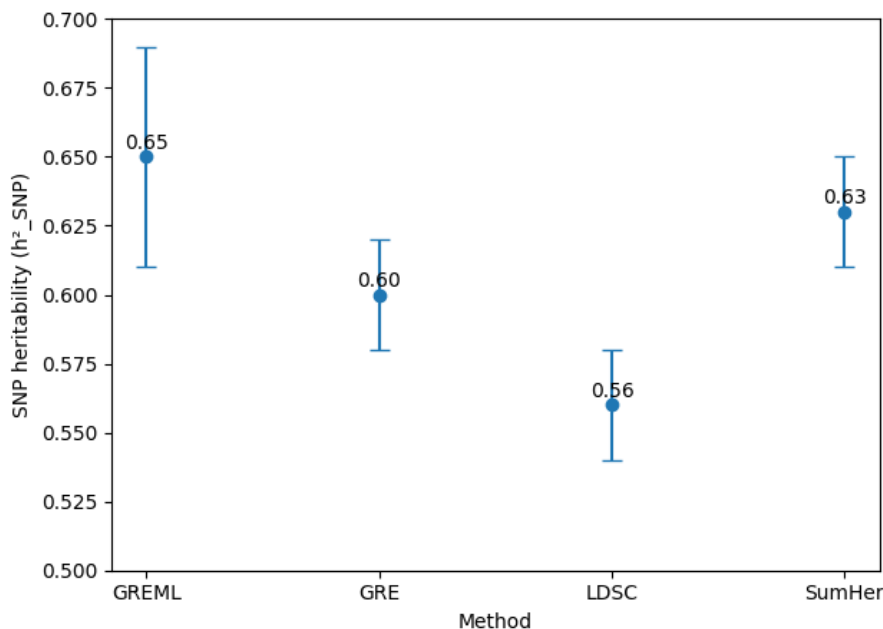


Figure 1 Cross-method comparison of SNP heritability estimates in UK Biobank height

Note: Bar plot showing SNP-based heritability estimates (h^2_{SNP}) across different methods. GREML-based approaches yield the highest estimates (~0.65), reflecting greater capture of genetic variance using individual-level data. LDSC produces systematically lower estimates (~0.56), likely due to reliance on summary statistics and LD reference assumptions. SumHer provides intermediate estimates (~0.63), incorporating LD- and MAF-dependent genetic architecture. The systematic differences illustrate method-dependent biases and support the concept of estimand mismatch