Moreover, when LD between causal variants and genotyped markers is weak, the effects of causal loci cannot be fully captured by tagging SNPs, leading to systematic underestimation of SNP-based heritability (Speed et al., 2012; 2016; Evans et al., 2017). The relationship between effect size distribution and the AF spectrum is also critical: when genetic contributions are driven primarily by rare variants or variants located in low-LD regions, the discrepancy between SNP-based and pedigree-based heritability is further amplified (Speed et al., 2016; Evans et al., 2017; Wainschtein et al., 2022).

6.1.1 Necessary conditions for comparing "heritability differences": from phenomenon to statistical framework

When discussing the discrepancy between pedigree-based heritability and SNP-based heritability, a frequently overlooked yet fundamental issue is whether such a comparison is statistically valid in the first place. These two types of estimates arise from distinct data structures and modeling frameworks; their differences are therefore not merely numerical deviations, but are embedded within their respective variance decomposition systems. Without strict alignment of underlying assumptions and conditions, the so-called "difference" often reflects only a superficial contrast between heterogeneous statistical objects, rather than an interpretable biological signal.

Consistency in phenotype definition constitutes the foundation of any meaningful comparison. A phenotype is not simply an observed variable; it directly embodies the variance structure subject to decomposition. Differences in measurement protocols, normalization procedures, or aggregation strategies across time points or traits can all alter the composition of phenotypic variance, thereby affecting both the numerator and denominator of heritability estimates. Once the phenotype definition shifts, even identical underlying genetic effects may yield systematically different estimates. As a result, comparisons lacking a unified phenotypic framework are unlikely to possess statistical interpretability.

The distribution of environmental factors and the structure of measurement error further define the reference frame for heritability estimation. Heritability is, by definition, the proportion of genetic variance relative to total phenotypic variance, and the environmental contribution to this total is highly dependent on the population context and study design. If studies differ substantially in environmental exposure, population composition, or sources of error, the decomposed variance components no longer belong to a common statistical population. Under such conditions, comparisons of heritability lose their foundation in a shared probability space.

In pedigree-based models, the treatment of shared environmental effects plays a critical role in identifying genetic variance. In twin or family studies, phenotypic similarity among related individuals arises from both genetic and shared environmental sources. If the model fails to adequately disentangle these components, part of the environmental effect may be misattributed to genetic variance, leading to systematic overestimation of pedigree heritability. This bias is structural rather than random, and often manifests as an apparent inflation of pedigree-based estimates relative to SNP-based heritability.

Finally, the coverage of the SNP marker system imposes a fundamental constraint on SNP-based heritability estimates. Estimates derived from genotyping arrays or sequencing data can only capture the genetic variation represented by observed markers and their linkage disequilibrium with causal variants. If low-frequency variants, rare variants, or structural variants are insufficiently represented, the corresponding genetic variance will be systematically missed. Therefore, even under correct model specification, SNP-based heritability cannot, in principle, reach the total level reflected by pedigree-based estimates.

6.1.2 Conceptual boundaries and interpretation of snp-based heritability

Within this analytical framework, the concept of SNP-based heritability requires a more precise definition. Rather than viewing it as a "lower-bound estimate" or a proxy for total trait heritability, it is more appropriately understood as the proportion of genetic variance captured by the observed SNP set under specific marker coverage and modeling assumptions-commonly referred to as "chip-capturable heritability." This definition highlights its conditional and tool-dependent nature, rather than treating it as a comprehensive representation of the genetic architecture of a trait.