Article

SumVg: Total Heritability Explained by All Variants in Genome-Wide Association Studies Based on Summary Statistics with Standard Error Estimates

Hon-Cheong So [1,2,3,4,5,6,7,*] , Xiao Xue [1], Zhijie Ma [1] and Pak-Chung Sham [8]

1   School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 1155151623@link.cuhk.edu.hk (X.X.); zhijiema@link.cuhk.edu.hk (Z.M.)
2   KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research of Common Diseases, Kunming Institute of Zoology and The Chinese University of Hong Kong, Shatin, Hong Kong, China
3   Department of Psychiatry, The Chinese University of Hong Kong, Shatin, Hong Kong, China
4   CUHK Shenzhen Research Institute, Shenzhen 518057, China
5   Margaret K. L. Cheung Research Centre for Management of Parkinsonism, The Chinese University of Hong Kong, Shatin, Hong Kong, China
6   Hong Kong Branch of the Chinese Academy of Sciences Center for Excellence in Animal Evolution and Genetics, The Chinese University of Hong Kong, Shatin, Hong Kong, China
7   Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, Hong Kong, China
8   Department of Psychiatry, The University of Hong Kong, Pokfulam, Hong Kong, China; pcsham@hku.hk
*   Correspondence: hcso@cuhk.edu.hk

Abstract: Genome-wide association studies (GWAS) are commonly employed to study the genetic basis of complex traits/diseases, and a key question is how much heritability could be explained by all single nucleotide polymorphisms (SNPs) in GWAS. One widely used approach that relies on summary statistics only is linkage disequilibrium score regression (LDSC); however, this approach requires certain assumptions about the effects of SNPs (e.g., all SNPs contribute to heritability and each SNP contributes equal variance). More flexible modeling methods may be useful. We previously developed an approach recovering the "true" effect sizes from a set of observed z-statistics with an empirical Bayes approach, using only summary statistics. However, methods for standard error (SE) estimation are not available yet, limiting the interpretation of our results and the applicability of the approach. In this study, we developed several resampling-based approaches to estimate the SE of SNP-based heritability, including two jackknife and three parametric bootstrap methods. The resampling procedures are performed at the SNP level as it is most common to estimate heritability from GWAS summary statistics alone. Simulations showed that the delete-d-jackknife and parametric bootstrap approaches provide good estimates of the SE. In particular, the parametric bootstrap approaches yield the lowest root-mean-squared-error (RMSE) of the true SE. We also explored various methods for constructing confidence intervals (CIs). In addition, we applied our method to estimate the SNP-based heritability of 12 immune-related traits (levels of cytokines and growth factors) to shed light on their genetic architecture. We also implemented the methods to compute the sum of heritability explained and the corresponding SE in an R package SumVg. In conclusion, SumVg may provide a useful alternative tool for calculating SNP heritability and estimating SE/CI, which does not rely on distributional assumptions of SNP effects.

Keywords: genome-wide association studies; SNP heritability; genetic epidemiology; bioinformatics; immunogenetics

1. Introduction

Genome-wide association studies (GWAS) have proven to be successful in dissecting the genetic basis of a variety of diseases. A number of new susceptibility loci have been discovered, providing novel insight into the pathophysiology of many diseases.

Nevertheless, a large proportion of heritability still remains unexplained. It is natural to question the maximum variance that could be explained by all variants in a GWAS (or meta-analyses of GWAS), as we expect that many true susceptibility variants are "hidden" due to limited power.

A number of methods have been developed to estimate total heritability according to all measured SNPs (also known as SNP-based heritability). Regarding methods that require individual-level data, in a pioneering work, Yang et al. [1] derived a method to estimate the variance explained by all SNPs in a GWAS using a linear mixed model with random SNP effects. The approach assumes that all SNPs have non-zero and normally distributed effects (beta), with a mean effect of zero. Each SNP is assumed to contribute to the same level of explained variance (i.e., variance explained by each SNP = total heritability/number of SNPs). Other similar approaches have also been proposed. For example, LDAK [2] assumes that a different heritability explains each SNP, depending on the minor allele frequencies (MAF), linkage disequilibrium (LD) score and imputation quality of the SNP. Advanced methods have also been developed to estimate SNP-based heritability using summary statistics alone. (Here, summary statistics refer to GWAS results for each SNP, with effect size (beta), standard error of beta and test statistics/p-values available, or at least two items available.) LD score regression (LDSC) is one of the most widely used approaches for this purpose [3]. LDSC assumes a mean effect (beta) of zero and equal variance explained by each SNP (i.e., an infinitesimal model). SumHer [4] is an alternative approach based on the LDAK assumptions. For a more detailed technical review, please refer to ref [5]. The broader problem of SNP-based heritability estimation has also been discussed in several other reviews or opinion pieces [6–9].

Prior to the development of LDSC, we have developed an alternative framework (see ref [10]; referred to as "SumVg" in this paper) to achieve the same goal of estimating SNP-based heritability using summary statistics alone. Essentially, we aimed to recover the *true* effect sizes from a set of *observed* z-statistics based on formulas presented by Robbins [11] (who attributed the idea to Maurice Kenneth Tweedie), Brown [12] and Efron [13]. The corrected z-statistics are then converted to variance explained. There are several advantages to this method. Most importantly, the SumVg approach does *not* rely on any distributional assumptions of the effect sizes of susceptibility variants. In addition, it does *not* assume an equal amount of heritability is explained by each SNP, or that all SNPs contribute to the heritability (infinitesimal model). There are also no assumptions about the relationship between allele frequencies and variance explained. The method is also computationally fast. In addition, since the LDSC method directly leverages LD patterns, a well-matched LD reference panel is usually required [14]. There is less reliance on LD information when using SumVg as LD is mainly used for pruning.

Our method has been applied in a number of studies (for example see [15–23]). However, there are no methods available to quantify the standard error (SE) or precision of the heritability estimates from SumVg, or the corresponding confidence intervals (CIs). There is considerable technical difficulty in developing a reliable approach for estimating the SE since usually only the summary data (instead of individual-level data) are available. If raw data are available, a standard non-parametric bootstrap could be employed by sampling individuals with a replacement. However, there are currently no methods for evaluating the SE or CI of the point estimate of heritability when only summary statistics are available.

We summarize the contributions of this study below. In this work, we proposed five re-sampling approaches to estimate the SE of the total heritability of all SNPs in GWAS, based on summary statistics alone. Extensive simulations were performed to compare and validate the performance of different methods. We also explored various methods for constructing CIs. Secondly, we also developed an easy-to-use R program to implement the SumVg approach with different flexible modeling options, available at https://github.com/lab-hcso/Estimating-SE-of-total-heritability/ (accessed on 12 October 2023). Thirdly, we reported heritability estimates for 12 immune-related traits (levels of cytokines and growth factors) [24] based on this approach, for which LDSC was unable to

provide reasonable estimates. Such cytokines/growth factors are regulators of immune responses and inflammation, and are important intermediate phenotypes for autoimmune, inflammatory and infectious diseases [25]. As such, it is of scientific and clinical importance to unravel the genetic architecture of these traits, and estimating their heritability may be considered a useful contribution in its own right.

2. Results

2.1. Overview of Methods

We estimated the total heritability (Vg) explained by all variants in a GWAS panel using the Tweedie's formula [10], which corrects selection bias in the observed z-statistics. To estimate the standard errors (SE), we proposed five resampling methods. The first two are based on jackknife, namely delete-one and delete-d-jackknife (with $d = n/5$ observations removed each time). We also proposed three parametric bootstrap methods, where z-statistics were sampled from a normal distribution based on the 'corrected' z-statistics, and/or the local false discovery rate (fdr) (i.e., estimated probability that a SNP is null). We also proposed several methods for constructing confidence intervals (CIs), including normal approximation with various bootstrap bias corrections, as well as the percentile and union of CI methods. We tested the performance of the SE and CI estimation methods in simulations under different heritability and sample size scenarios. We applied our methods to estimate SNP-based heritability and the SEs of 12 immune traits to reveal their genetic architecture.

2.2. Simulation Results for SE Estimation

Standard errors (SEs) of heritability, as estimated by the jackknife and bootstrap approaches, are listed in Table 1 and plotted in Figure 1. Bias, variance and root mean square error (RMSE) of SEs were calculated over 100 simulations (Table 2; Figures 1 and 2).

Table 1. Standard error (SE) of the sum of variance explained (Vg) estimated by different resampling approaches.

Sum_of_Vg	Sample_Size	Mean_Est	TRUE_SE	SE				
				jack_del_1	jack_del_d	paraboot	fdrboot1	fdrboot2
0.295	5000	0.232	0.0482	0.0672	0.0524	0.0488	0.0519	0.0489
	10,000	0.210	0.0265	0.0353	0.0295	0.0285	0.0312	0.0287
	20,000	0.244	0.0158	0.0208	0.0185	0.0165	0.0156	0.0168
	50,000	0.283	0.0076	0.0149	0.0167	0.0063	0.0081	0.0063
	1×10^5	0.312	0.0063	0.0143	0.0172	0.0051	0.0055	0.0054
	2×10^5	0.321	0.0045	0.0134	0.0161	0.0036	0.0038	0.0041
0.191	5000	0.207	0.0491	0.0706	0.0523	0.0486	0.0500	0.0485
	10,000	0.147	0.0242	0.0357	0.0274	0.0263	0.0285	0.0265
	20,000	0.158	0.0159	0.0208	0.0166	0.0156	0.0162	0.0160
	50,000	0.174	0.0064	0.0113	0.0113	0.0061	0.0070	0.0061
	1×10^5	0.195	0.0045	0.0110	0.0131	0.0040	0.0047	0.0041
	2×10^5	0.207	0.0035	0.0103	0.0128	0.0031	0.0034	0.0035
0.101	5000	0.197	0.0521	0.0692	0.0524	0.0484	0.0496	0.0483
	10,000	0.116	0.0260	0.0345	0.0265	0.0251	0.0257	0.0251
	20,000	0.098	0.0143	0.0202	0.0159	0.0150	0.0153	0.0154
	50,000	0.091	0.0058	0.0098	0.0078	0.0063	0.0057	0.0063
	1×10^5	0.094	0.0032	0.0069	0.0076	0.0027	0.0036	0.0027
	2×10^5	0.107	0.0028	0.0072	0.0083	0.0023	0.0027	0.0025

The main purpose of this table is to compare the true SEs against the SEs estimated by various resampling-based methods. Sum_of_Vg, true total heritability explained (i.e., the real total heritability based on our data-generating mechanism); Sample_size refers to the sample size of the GWAS; Mean_Est, the mean *estimated* heritability explained based on our approach of corrected z-statistics; True_SE, the 'true' SE was based on repeating our simulation experiments 100 times; jack_del_1, delete-1-jackknife; jack_del_d, delete-d-jackknife with d equal to 20,000; paraboot, parametric bootstrap approach as described in the text, based on simulating from a normal distribution in which the mean was derived from the corrected z-statistics (without consideration of local false discovery rates (fdr)); fdrboot1, a "weighted" bootstrap approach with consideration of the local fdr, using the observed z-statistic as the mean in each simulation; fdrboot2, a "weighted" bootstrap approach with consideration of the local fdr, using the corrected z-statistic as the mean in each simulation.
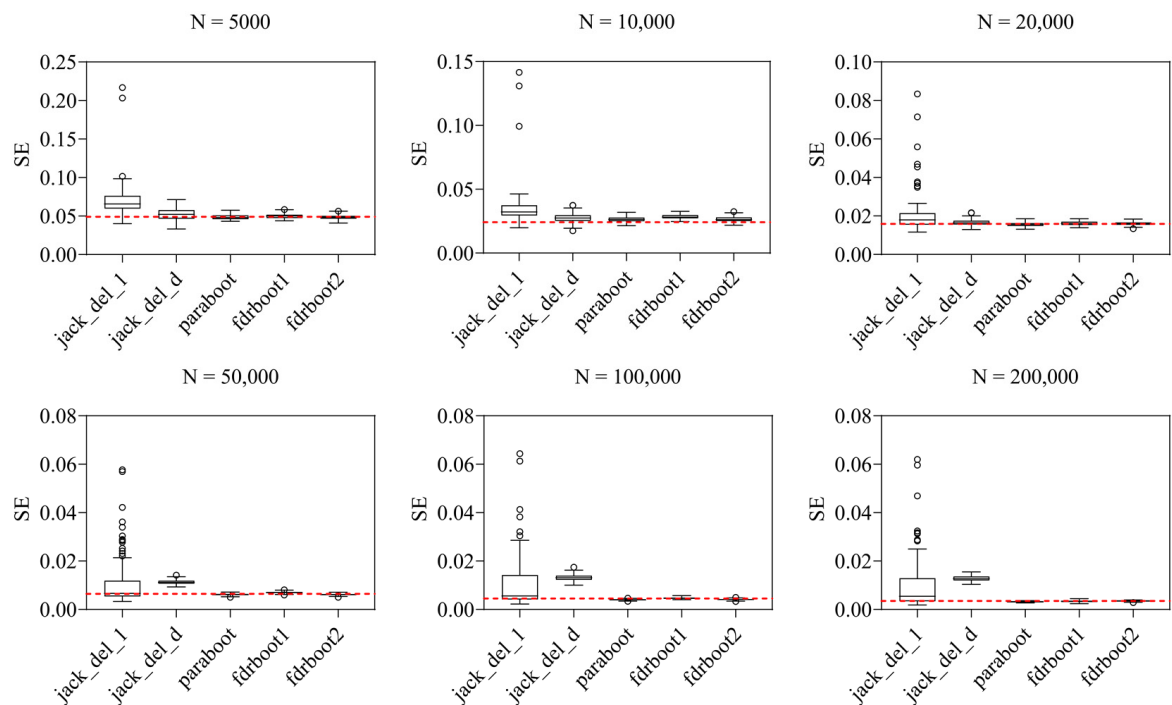
Vg=0.295
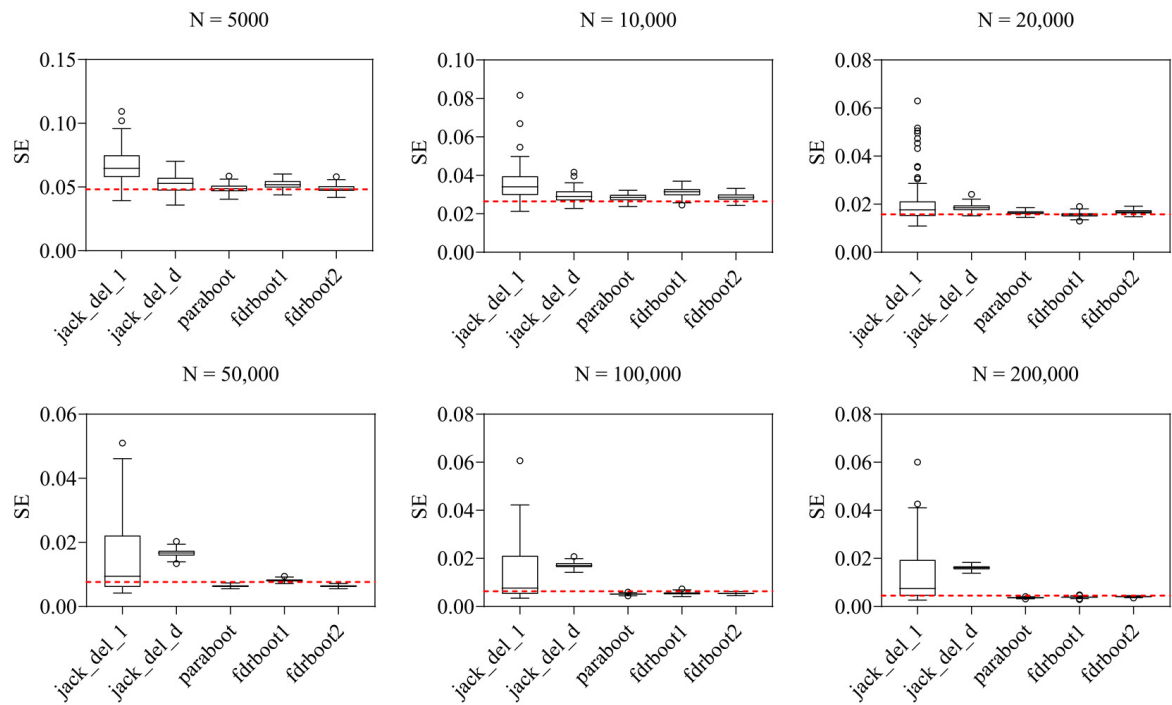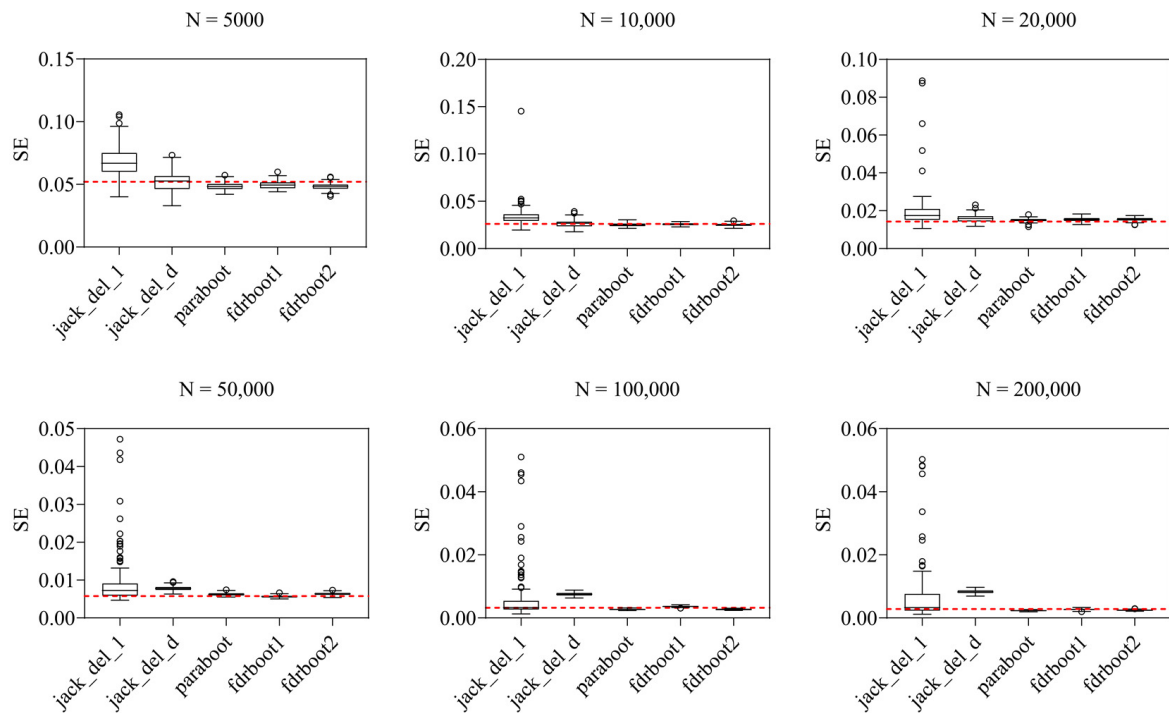

Vg=0.191

Figure 1. *Cont.*

Vg=0.101

Figure 1. Boxplots of SE estimated by different approaches. Vg is the sum of variance explained, *N* is the sample size, and the horizontal line refers to true SE calculated by repeating the experiments 100 times based on the true data generating mechanism. jack_del_1, jack_del_d, paraboot, fdrboot1 and fdrboot2 are different SE estimation approaches as described above.

The delete-[$n/5$]-jackknife worked reasonably well when the total heritability explained was low (when heritability = 0.101), but it tended to overestimate the SE when the total heritability was higher, especially with larger sample sizes. The bias was also positive across all simulation scenarios. The standard (delete-1) jackknife approach performed the worst among all methods, producing inflated estimates of SE. The variance and RMSE of this estimator were high compared to other approaches. The SE was, in general, over-estimated at all heritability levels across all sample sizes. This may be explained by the fact that the sum of variance explained is not a very smooth parameter, which impairs the performance of delete-1-jackknife estimators.

The other methods, including the original parametric bootstrap (paraboot) and the modified versions with consideration of local fdr, performed reasonably well and closely resembled the true SE. With the exception of one simulation setting, the parametric bootstrap methods achieved the lowest (absolute) bias for SE. For the variance and RMSE of SE, parametric bootstrap also performed the best. In terms of RMSE, the parametric bootstrap approaches modeling the local fdr (i.e., fdrboot1 and fdrboot2) outperformed the other methods. The RMSE of different estimators were also observed to reduce with increasing sample sizes.

Table 2. Bias, variance and root mean squared error (RMSE) of SE estimated by different resampling approaches.

| Sum_Vg	N	Bias of the Estimator for SE					Variance of the Estimator for SE					RMSE of the Estimator for SE				
		jack_del_1	jack_del_d	paraboot	fdrboot1	fdrboot2	jack_del_1	jack_del_d	paraboot	fdrboot1	fdrboot2	jack_del_1	jack_del_d	paraboot	fdrboot1	fdrboot2
0.295	5000	1.91×10^{-2}	4.26×10^{-3}	$\mathbf{5.92 \times 10^{-4}}$	3.73×10^{-3}	7.16×10^{-4}	1.77×10^{-4}	5.14×10^{-5}	1.26×10^{-5}	1.15×10^{-5}	$\mathbf{9.30 \times 10^{-6}}$	2.32×10^{-2}	8.34×10^{-3}	3.59×10^{-3}	5.04×10^{-3}	$\mathbf{3.13 \times 10^{-3}}$
	10,000	8.81×10^{-3}	2.98×10^{-3}	$\mathbf{2.00 \times 10^{-3}}$	4.66×10^{-3}	2.25×10^{-3}	7.34×10^{-5}	1.21×10^{-5}	3.87×10^{-6}	4.99×10^{-6}	$\mathbf{3.62 \times 10^{-6}}$	1.23×10^{-2}	4.58×10^{-3}	$\mathbf{2.80 \times 10^{-3}}$	5.17×10^{-3}	2.94×10^{-3}
	20,000	5.04×10^{-3}	2.78×10^{-3}	7.37×10^{-4}	$\mathbf{-1.46 \times 10^{-4}}$	1.07×10^{-3}	9.06×10^{-5}	2.27×10^{-6}	$\mathbf{8.33 \times 10^{-7}}$	1.12×10^{-6}	1.00×10^{-6}	1.08×10^{-2}	3.16×10^{-3}	1.17×10^{-3}	$\mathbf{1.07 \times 10^{-3}}$	1.47×10^{-3}
	50,000	7.25×10^{-3}	9.03×10^{-3}	-1.32×10^{-3}	$\mathbf{4.68 \times 10^{-4}}$	-1.30×10^{-3}	1.29×10^{-4}	1.45×10^{-6}	1.36×10^{-7}	1.93×10^{-7}	$\mathbf{1.30 \times 10^{-7}}$	1.35×10^{-2}	9.11×10^{-3}	1.37×10^{-3}	$\mathbf{6.42 \times 10^{-4}}$	1.35×10^{-3}
	1×10^{5}	7.97×10^{-3}	1.09×10^{-2}	-1.20×10^{-3}	$\mathbf{-8.78 \times 10^{-4}}$	-9.34×10^{-4}	1.41×10^{-4}	1.52×10^{-6}	$\mathbf{8.11 \times 10^{-8}}$	3.48×10^{-7}	1.00×10^{-7}	1.43×10^{-2}	1.10×10^{-2}	1.23×10^{-3}	1.06×10^{-3}	$\mathbf{9.86 \times 10^{-4}}$
	2×10^{5}	8.92×10^{-3}	1.16×10^{-2}	-8.57×10^{-4}	-6.32×10^{-4}	$\mathbf{-3.72 \times 10^{-4}}$	1.37×10^{-4}	8.70×10^{-7}	$\mathbf{3.49 \times 10^{-8}}$	1.30×10^{-7}	4.06×10^{-8}	1.47×10^{-2}	1.16×10^{-2}	8.77×10^{-4}	7.27×10^{-4}	$\mathbf{4.23 \times 10^{-4}}$
0.191	5000	2.16×10^{-2}	3.21×10^{-3}	$\mathbf{-5.02 \times 10^{-4}}$	9.69×10^{-4}	-5.23×10^{-4}	5.53×10^{-4}	5.41×10^{-5}	1.07×10^{-5}	1.02×10^{-5}	$\mathbf{8.58 \times 10^{-6}}$	3.19×10^{-2}	8.03×10^{-3}	3.31×10^{-3}	3.34×10^{-3}	$\mathbf{2.98 \times 10^{-3}}$
	10,000	1.15×10^{-2}	3.16×10^{-3}	$\mathbf{2.04 \times 10^{-3}}$	4.22×10^{-3}	2.29×10^{-3}	2.80×10^{-4}	1.43×10^{-5}	4.88×10^{-6}	$\mathbf{2.98 \times 10^{-6}}$	5.03×10^{-6}	2.03×10^{-2}	4.93×10^{-3}	$\mathbf{3.01 \times 10^{-3}}$	4.56×10^{-3}	3.20×10^{-3}
	20,000	4.96×10^{-3}	7.22×10^{-4}	-2.41×10^{-4}	3.15×10^{-4}	$\mathbf{9.85 \times 10^{-5}}$	1.19×10^{-4}	2.70×10^{-6}	1.09×10^{-6}	1.31×10^{-6}	$\mathbf{7.64 \times 10^{-7}}$	1.20×10^{-2}	1.79×10^{-3}	1.07×10^{-3}	1.19×10^{-3}	$\mathbf{8.80 \times 10^{-4}}$
	50,000	4.90×10^{-3}	4.83×10^{-3}	$\mathbf{-2.92 \times 10^{-4}}$	5.76×10^{-4}	-2.97×10^{-4}	1.10×10^{-4}	8.17×10^{-7}	1.66×10^{-7}	1.50×10^{-7}	$\mathbf{1.24 \times 10^{-7}}$	1.16×10^{-2}	4.91×10^{-3}	5.01×10^{-4}	6.94×10^{-4}	$\mathbf{4.60 \times 10^{-4}}$
	1×10^{5}	6.45×10^{-3}	8.56×10^{-3}	-5.40×10^{-4}	$\mathbf{1.30 \times 10^{-4}}$	-4.33×10^{-4}	1.32×10^{-4}	1.68×10^{-6}	$\mathbf{6.30 \times 10^{-8}}$	1.32×10^{-7}	8.37×10^{-8}	1.32×10^{-2}	8.65×10^{-3}	5.96×10^{-4}	$\mathbf{3.85 \times 10^{-4}}$	5.21×10^{-4}
	2×10^{5}	6.85×10^{-3}	9.28×10^{-3}	-3.57×10^{-4}	-1.41×10^{-4}	$\mathbf{-3.31 \times 10^{-5}}$	1.28×10^{-4}	1.06×10^{-6}	$\mathbf{3.04 \times 10^{-8}}$	1.54×10^{-7}	3.78×10^{-8}	1.32×10^{-2}	9.34×10^{-3}	3.97×10^{-4}	4.17×10^{-4}	$\mathbf{1.97 \times 10^{-4}}$
0.101	5000	1.71×10^{-2}	$\mathbf{3.38 \times 10^{-4}}$	-3.72×10^{-3}	-2.54×10^{-3}	-3.81×10^{-3}	1.81×10^{-4}	6.26×10^{-5}	1.01×10^{-5}	1.07×10^{-5}	$\mathbf{7.70 \times 10^{-6}}$	2.17×10^{-2}	7.92×10^{-3}	4.90×10^{-3}	$\mathbf{4.14 \times 10^{-3}}$	4.71×10^{-3}
	10,000	8.45×10^{-3}	4.62×10^{-4}	-8.81×10^{-4}	$\mathbf{-2.84 \times 10^{-4}}$	-9.03×10^{-4}	1.66×10^{-4}	1.45×10^{-5}	3.49×10^{-6}	$\mathbf{1.66 \times 10^{-6}}$	2.46×10^{-6}	1.54×10^{-2}	3.83×10^{-3}	2.07×10^{-3}	$\mathbf{1.32 \times 10^{-3}}$	1.81×10^{-3}
	20,000	5.92×10^{-3}	1.56×10^{-3}	$\mathbf{7.21 \times 10^{-4}}$	1.04×10^{-3}	1.08×10^{-3}	1.44×10^{-4}	4.20×10^{-6}	8.80×10^{-7}	1.22×10^{-6}	$\mathbf{8.17 \times 10^{-7}}$	1.34×10^{-2}	2.57×10^{-3}	$\mathbf{1.18 \times 10^{-3}}$	1.52×10^{-3}	1.41×10^{-3}
	50,000	4.03×10^{-3}	2.04×10^{-3}	4.91×10^{-4}	$\mathbf{-1.02 \times 10^{-4}}$	5.85×10^{-4}	6.00×10^{-5}	4.31×10^{-7}	1.61×10^{-7}	$\mathbf{1.08 \times 10^{-7}}$	1.26×10^{-7}	8.73×10^{-3}	2.15×10^{-3}	6.34×10^{-4}	$\mathbf{3.44 \times 10^{-4}}$	6.85×10^{-4}
	1×10^{5}	3.73×10^{-3}	4.34×10^{-3}	-5.10×10^{-4}	$\mathbf{4.31 \times 10^{-4}}$	-5.17×10^{-4}	9.13×10^{-5}	3.03×10^{-7}	2.83×10^{-8}	4.34×10^{-8}	$\mathbf{2.05 \times 10^{-8}}$	1.03×10^{-2}	4.38×10^{-3}	5.37×10^{-4}	$\mathbf{4.79 \times 10^{-4}}$	5.36×10^{-4}
	2×10^{5}	4.38×10^{-3}	5.48×10^{-3}	-5.31×10^{-4}	$\mathbf{-1.61 \times 10^{-4}}$	-3.81×10^{-4}	1.00×10^{-4}	3.41×10^{-7}	$\mathbf{2.22 \times 10^{-8}}$	7.06×10^{-8}	2.88×10^{-8}	1.09×10^{-2}	5.51×10^{-3}	5.52×10^{-4}	$\mathbf{3.11 \times 10^{-4}}$	4.17×10^{-4}

The table shows the bias, variance and root mean squared error of SE estimated from our methods, as compared to the true SE. Sum_Vg, true total heritability explained; *N*, sample size. The best performing method (for estimation of SE) in each scenario is in bold. For other abbreviations, please refer to Table 1.
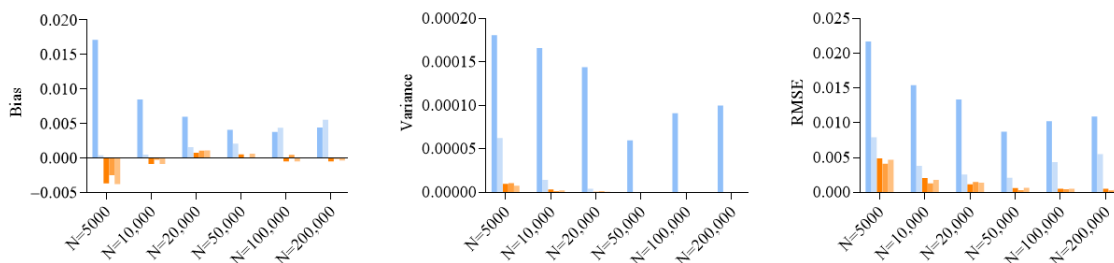
Vg=0.295
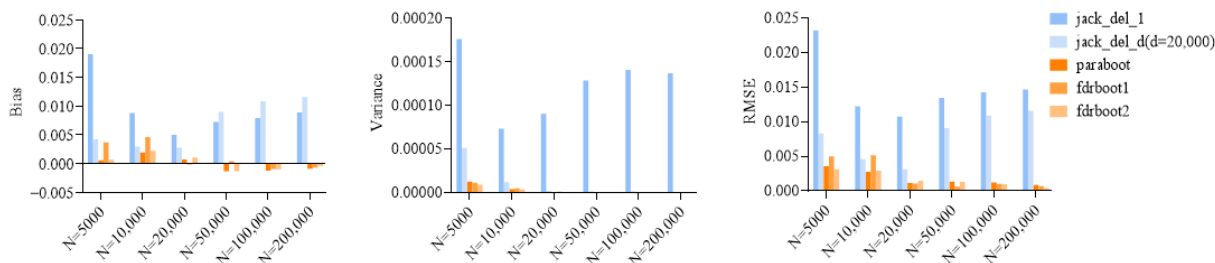
Vg=0.191
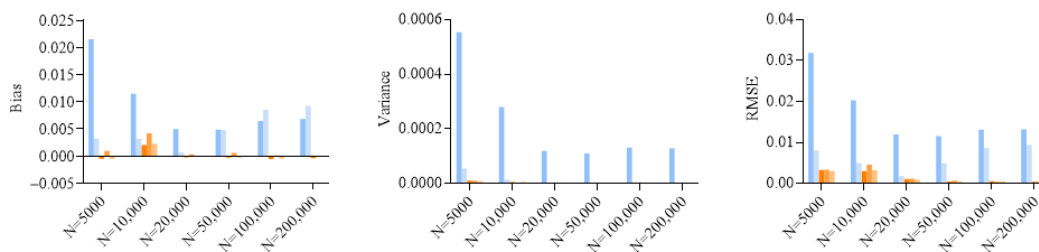
Vg=0.101

Figure 2. Bar plots of bias, variance and root mean squared error (RMSE) of SE estimated by different approaches in simulations. Vg is the sum of variance explained, and N is the sample size. jack_del_1, jack_del_d, paraboot, fdrboot1 and fdrboot2 are different SE estimation approaches as described above.

2.3. Performance of Different CI Construction Methods

The full results are presented in Tables 3, S1 and S2. For standard CI (based on normal approximation), the CIs built from the SE of delete-*d*-jackknife performed reasonably well (in terms of coverage) for large sample sizes, although the coverage was not always adequate for modest samples sizes, especially for *N* < 20,000. The coverage of CIs constructed from other types of SEs were more variable, with good coverage for some scenarios but poor coverage for others. Therefore, we primarily focus on the SE from delete-*d*-jackknife when a standard CI is used. Interestingly, the bias-corrected standard CI, with bias correction based on paraboot or fdrboot2, performed better in the several cases when the standard CI had low coverage (<50%) (we assume that the SE from delete-*d*-jackknife was employed). The performance of percentile CIs was highly variable across different scenarios.

Table 3. Coverage probabilities of different union CI (UCI) approaches for 95% CI.

N	Union CI Type	Coverage (Vg = 0.295)	Coverage (Vg = 0.191)	Coverage (Vg = 0.101)
5000	Standard	0.75	0.97	0.77
	Percentile	1	1	0.78
	Standard + Percentile	1	1	0.78
10,000	Standard	0.6	0.67	0.94
	Percentile	0.99	1	1
	Standard + Percentile	0.99	1	1
20,000	Standard	0.89	0.84	0.96
	Percentile	0.91	1	1
	Standard + Percentile	0.97	1	1
50,000	Standard	1	1	0.9
	Percentile	1	1	1
	Standard + Percentile	1	1	1
1×10^5	Standard	1	1	1
	Percentile	1	1	1
	Standard + Percentile	1	1	1
2×10^5	Standard	0.96	1	1
	Percentile	0.13	0.66	1
	Standard+Percentile	0.96	1	1

Notes for Table 3: The following methods were explored: 1. Normal approximation (standard approach) without bias correction (one estimator) or with bootstrap bias correction (3 estimators), and the union of CIs was taken; ("Standard"). 2. Percentile approach without bias correction (3 estimators) and with bias correction (3 estimators), and the union of CIs was taken; ("Percentile"). 3. Union of the final union CIs obtained from 1 and 2. ("Standard+Percentile"). Coverage refers to the coverage probabilities based on simulations.

In view of the highly variable performance of different CI construction methods, we expect the union of CI (UCI) to perform better and be more robust across different scenarios. We observed that UCI, no matter if it is constructed from the standard or percentile CI estimators, in general, achieved good coverage across most simulation scenarios, although in some cases the coverage was still below the desired level (95%). When we further took the union of standard and percentile UCI estimators (i.e., Method 3 listed under 'Union CI' in the Section 4), the coverage was adequate for almost all scenarios, except one case in which both the sample size and the sum of variance explained (Vg) were low (*N* = 5000, Vg = 0.101).

2.4. Results on Immune Traits

PLINK was applied to trim GWAS data for 12 immunological traits (Table 4) with various r^2 criteria to obtain roughly independent SNPs. We only included common variants with an MAF > 0.01 for further analysis. Then, using SumVg, the "true" *z*-statistics of trimmed SNPs were retrieved to capture the missing heritability. The jackknife and bootstrap methods were used to compute the corresponding SEs (Table 5; Figure S1).

The total SNP-based heritability predicted by SumVg for the selected traits, in contrast to the comparatively low or negative heritability estimates from LDSC, were around 10–20% based on a collection of LD-pruned SNPs. We obtained a stable (and likely conservative) estimate of heritability at $r^2 \sim 0.01$ or 0.005. Lower r^2 values (i.e., $r^2 < 0.0025$ and $r^2 < 0.001$) had limited impact on final estimates of heritability. The delete-one jackknife consistently produced the highest standard error, while the bootstrap and delete-*d* jackknife approaches produced SEs that were more comparable to one another. Out of the 12 cytokines/growth factors studied, the highest heritability was observed for the levels of IL-4 and IL-17.

Table 4. Summary of the immune traits being studied.

Trait	Abbreviation	GWAS ID	N	SNP_h2 (LDSC)	SNP_h2_se (LDSC)
Stem cell factor	SCF	ebi-a-GCST004429	8290	−0.06	0.055
Interleukin-4	IL4	ebi-a-GCST004453	8124	−0.0446	0.0595
Interleukin-17	IL17	ebi-a-GCST004442	7760	−0.0407	0.0623
Hepatocyte growth factor	HGF	ebi-a-GCST004449	8292	−0.0311	0.0579
Basic fibroblast growth factor	FGFBasic	ebi-a-GCST004459	7565	−0.0159	0.0597
Stromal cell-derived factor-1 alpha (CXCL12)	SDF1a	ebi-a-GCST004427	5998	−0.0116	0.0713
Interleukin-6	IL6	ebi-a-GCST004446	8189	−0.0071	0.0568
Platelet derived growth factor BB	PDGFbb	ebi-a-GCST004432	8293	−0.0043	0.0624
TNF-related apoptosis inducing ligand	TRAIL	ebi-a-GCST004424	8186	0.0125	0.0613
Interferon-gamma	IFNg	ebi-a-GCST004456	7701	0.0134	0.0624
Granulocyte colony-stimulating factor	GCSF	ebi-a-GCST004458	7904	0.0173	0.0601
Interleukin-10	IL10	ebi-a-GCST004444	7681	0.0186	0.0691

Trait, trait name of analyzed GWAS dataset; abbreviation, abbreviation of the trait name; GWAS ID, ID of GWAS dataset for downloading from the IEU OpenGWAS Project; N, sample size; SNP_h2 (LDSC), SNP heritability estimated by LDSC as reported in GWASAtlas; SNP_h2_se (LDSC), standard error of SNP heritability estimated by LDSC as reported in GWASAtlas.

Table 5. SE of the sum of variance explained estimated by different resampling approaches, for 12 immune traits (under different r^2 pruning thresholds).

Trait	N	LDSC h2	se	h2	r^2	n_pruned_snp	SumVg se_jack1	se_jack_del_d	se_paraboot	se_fdrboot1	se_fdrboot2
SCF	8290	−0.06	0.055	0.333	0.1	428,593	0.0926	0.0822	0.0679	0.0443	0.0514
				0.185	0.05	251,008	0.0526	0.0456	0.0467	0.0502	0.0517
				0.105	0.025	127,908	0.0307	0.0313	0.0272	0.0397	0.0335
				0.100	**0.01**	**61,938**	**0.0310**	**0.0200**	**0.0220**	**0.0252**	**0.0265**
				0.092	0.005	51,370	0.0229	0.0169	0.0201	0.0235	0.0230
				0.101	0.002	48,088	0.0319	0.0153	0.0220	0.0226	0.0198
				0.102	0.001	47,108	0.0316	0.0155	0.0223	0.0216	0.0188
IL4	8124	−0.0446	0.0595	0.503	0.1	427,005	0.1218	0.1133	0.0616	0.0563	0.0569
				0.377	0.05	249,710	0.1000	0.0823	0.0484	0.0445	0.0453
				0.302	0.025	127,248	0.0650	0.0594	0.0318	0.0365	0.0336
				0.235	**0.01**	**61,685**	**0.0529**	**0.0313**	**0.0247**	**0.0240**	**0.0236**
				0.215	0.005	51,196	0.0472	0.0278	0.0227	0.0217	0.0221
				0.197	0.002	47,878	0.0571	0.0273	0.0228	0.0225	0.0253
				0.187	0.001	46,911	0.0482	0.0244	0.0198	0.0226	0.0242
IL17	7760	−0.0407	0.0623	0.352	0.1	427,226	0.1240	0.0946	0.0692	0.0625	0.0609
				0.228	0.05	250,259	0.0683	0.0499	0.0360	0.0495	0.0495
				0.299	0.025	127,479	0.0877	0.0568	0.0360	0.0380	0.0323
				0.234	**0.01**	**61,756**	**0.0485**	**0.0340**	**0.0239**	**0.0267**	**0.0256**
				0.196	0.005	51,215	0.0475	0.0295	0.0237	0.0190	0.0249
				0.195	0.002	47,887	0.0634	0.0231	0.0231	0.0226	0.0210
				0.188	0.001	46,931	0.0568	0.0242	0.0183	0.0211	0.0215
HGF	8292	−0.0311	0.0579	0.366	0.1	428,318	0.0917	0.0864	0.0569	0.0642	0.0593
				0.242	0.05	250,843	0.0812	0.0722	0.0483	0.0492	0.0491
				0.205	0.025	127,850	0.0657	0.0488	0.0327	0.0326	0.0357
				0.098	**0.01**	**61,906**	**0.0379**	**0.0224**	**0.0225**	**0.0260**	**0.0242**
				0.115	0.005	51,301	0.0347	0.0199	0.0224	0.0230	0.0203
				0.111	0.002	47,878	0.0414	0.0162	0.0189	0.0211	0.0215
				0.108	0.001	46,934	0.0312	0.0171	0.0221	0.0208	0.0211
FGFBasic	7565	−0.0159	0.0597	0.269	0.1	427,284	0.0835	0.0902	0.0656	0.0530	0.0577
				0.217	0.05	249,930	0.0891	0.0604	0.0473	0.0504	0.0468
				0.117	0.025	127,587	0.0452	0.0431	0.0340	0.0363	0.0358
				0.133	**0.01**	**61,911**	**0.0408**	**0.0301**	**0.0232**	**0.0239**	**0.0275**
				0.135	0.005	51,259	0.0376	0.0243	0.0242	0.0267	0.0219
				0.143	0.002	47,874	0.0362	0.0218	0.0185	0.0233	0.0245
				0.126	0.001	46,914	0.0392	0.0206	0.0227	0.0214	0.0208
SDF1a	5998	−0.0116	0.0713	0.395	0.1	425,165	0.1120	0.1068	0.0731	0.0757	0.0870
				0.256	0.05	248,727	0.0872	0.0750	0.0580	0.0565	0.0631
				0.213	0.025	126,986	0.0707	0.0462	0.0431	0.0472	0.0468
				0.163	**0.01**	**61,680**	**0.0497**	**0.0380**	**0.0359**	**0.0349**	**0.0324**
				0.190	0.005	51,092	0.0708	0.0318	0.0250	0.0297	0.0270
				0.165	0.002	47,702	0.0447	0.0270	0.0294	0.0304	0.0301
				0.159	0.001	46,789	0.0512	0.0232	0.0279	0.0258	0.0308
IL6	8189	−0.0071	0.0568	0.422	0.1	427,566	0.0878	0.0896	0.0510	0.0575	0.0594
				0.227	0.05	250,247	0.0620	0.0713	0.0402	0.0463	0.0468
				0.158	0.025	127,503	0.0672	0.0457	0.0372	0.0300	0.0360
				0.139	**0.01**	**61,931**	**0.0606**	**0.0258**	**0.0220**	**0.0247**	**0.0220**
				0.114	0.005	51,332	0.0288	0.0176	0.0196	0.0227	0.0236
				0.115	0.002	47,930	0.0302	0.0164	0.0191	0.0226	0.0202
				0.117	0.001	46,944	0.0319	0.0175	0.0227	0.0209	0.0211

Table 5. *Cont.*

Trait	N	LDSC		SumVg							
		h2	se	h2	r^2	n_pruned_snp	se_jack1	se_jack_del_d	se_paraboot	se_fdrboot1	se_fdrboot2
PDGFbb	8293	−0.0043	0.0624	0.432	0.1	427,743	0.0907	0.0993	0.0726	0.0653	0.0676
				0.341	0.05	250,325	0.0670	0.0808	0.0600	0.0496	0.0576
				0.307	0.025	127,567	0.0735	0.0554	0.0370	0.0334	0.0326
				0.154	**0.01**	**61,789**	**0.0372**	**0.0250**	**0.0213**	**0.0245**	**0.0243**
				0.125	0.005	51,140	0.0310	0.0226	0.0234	0.0230	0.0221
				0.120	0.002	47,822	0.0258	0.0205	0.0214	0.0208	0.0233
				0.117	0.001	46,853	0.0392	0.0192	0.0201	0.0209	0.0226
TRAIL	8186	0.0125	0.0613	0.559	0.1	423,391	0.0613	0.1018	0.0785	0.0790	0.0750
				0.304	0.05	247,717	0.0543	0.1190	0.0526	0.0503	0.0439
				0.242	0.025	126,350	0.0607	0.0647	0.0321	0.0362	0.0370
				0.128	**0.01**	**61,114**	**0.0316**	**0.0251**	**0.0242**	**0.0229**	**0.0277**
				0.127	0.005	50,633	0.0298	0.0231	0.0255	0.0239	0.0268
				0.128	0.002	47,359	0.0332	0.0216	0.0233	0.0215	0.0266
				0.121	0.001	46,415	0.0358	0.0195	0.0222	0.0229	0.0256
IFNg	7701	0.0134	0.0624	0.393	0.1	426,740	0.0946	0.0811	0.0528	0.0590	0.0594
				0.241	0.05	249,818	0.0655	0.0628	0.0553	0.0520	0.0509
				0.244	0.025	127,514	0.0734	0.0582	0.0330	0.0406	0.0320
				0.138	**0.01**	**61,890**	**0.0289**	**0.0303**	**0.0267**	**0.0239**	**0.0257**
				0.138	0.005	51,314	0.0424	0.0201	0.0222	0.0248	0.0293
				0.141	0.002	47,918	0.0321	0.0204	0.0251	0.0248	0.0286
				0.137	0.001	46,934	0.0253	0.0183	0.0223	0.0246	0.0233
GCSF	7904	0.0173	0.0601	0.246	0.1	427,393	0.0707	0.0820	0.0620	0.0604	0.0580
				0.198	0.05	250,222	0.0636	0.0607	0.0402	0.0436	0.0486
				0.164	0.025	127,583	0.0501	0.0415	0.0302	0.0360	0.0327
				0.142	**0.01**	**61,846**	**0.0434**	**0.0257**	**0.0280**	**0.0239**	**0.0257**
				0.122	0.005	51,266	0.0379	0.0196	0.0205	0.0247	0.0238
				0.120	0.002	47,919	0.0413	0.0183	0.0219	0.0236	0.0201
				0.112	0.001	46,939	0.0312	0.0159	0.0234	0.0207	0.0202
IL10	7681	0.0186	0.0691	0.331	0.1	427,218	0.0621	0.1019	0.0584	NA	NA
				0.310	0.05	250,109	0.0670	0.0858	0.0448	NA	NA
				0.198	0.025	127,543	0.0566	0.0463	0.0356	0.0382	0.0406
				0.130	**0.01**	**61,944**	**0.0328**	**0.0225**	**0.0251**	**0.0268**	**0.0258**
				0.141	0.005	51,257	0.0400	0.0220	0.0237	0.0282	0.0238
				0.148	0.002	47,880	0.0433	0.0183	0.0204	0.0271	0.0231
				0.142	0.001	46,898	0.0317	0.0194	0.0219	0.0261	0.0228

This table shows the estimated total SNP-based heritability and their SEs for 12 immune traits. We also show a comparison of the estimates between LDSC and SumVg. Trait, N, LDSC (h2, se) have the same meaning as in Table 4; h2, heritability estimated by SumVg across a set of r^2 pruning thresholds; r^2, the r^2 pruning threshold; n_pruned_snp, number of SNPs after LD pruning at the corresponding r^2 threshold; se_jack_1, se_jack_del_d, se_paraboot, se_fdrboot1 and se_fdrboot2 are SE estimated by different approaches as described above; "NA" was shown when "locfdr" failed to estimate local false discovery rate. The estimates with $r^2 = 0.01$ were highlighted, as we observed that in general the heritability estimates stabilize at $r^2 \sim 0.01$.

2.5. R Package Implementation

We also implemented the methods to compute the sum of heritability explained and the corresponding SEs in an R package SumVg, available at https://github.com/lab-hcso/Estimating-SE-of-total-heritability/ (accessed on 12 October 2023).

The computational speed of different resampling approaches using SumVg is presented in Table S3 (assuming 100,000 SNPs and 200 resampling iterations). The speed is generally fast and the time taken was around 2–4 min for each resampling method, using a single core (Intel Xeon Gold 6230 CPU @ 2.10 GHz).

3. Discussion

In this study, we presented an approach for estimating the SE of SNP-based heritability estimates using SumVg, and our applications to immune phenotypes demonstrate the usefulness of this approach.

Our main purpose is to provide an alternative approach for SNP-based heritability and SE estimation, since different approaches have different statistical modeling assumptions, or assumptions about the genetic architecture. In practice, it is almost impossible to know the true genetic architecture of a disease/trait, and as such, it is very difficult to verify the correctness of heritability estimates due to the lack of a 'gold standard'. It will be more reassuring if one observes similar heritability estimates from diverse methods. SumVg may provide a useful alternative reference for heritability estimates, in conjunction with existing approaches such as LDSC. SumVg may also be useful when standard approaches are unable to give reasonable results (e.g., close to zero heritability for traits that are likely to be heritable from previous studies, or negative estimates). It will be interesting to

investigate the reasons underlying negative heritability estimates for LDSC; one possibility is mis-specified model assumptions [26], but the exact reasons will require further studies.

We recommended pruning the SNPs (such that SNPs are roughly in linkage equilibrium) before applying our method of heritability estimation. One approach is to employ a series of r^2 thresholds (e.g., decreasing r^2 from 0.1 to 0.001) and consider the point at which heritability became stable. Our empirical applications showed that an r^2 threshold of ~0.01 may be sufficient. The resulting SNP-based heritability may be considered to be a conservative estimate (due to the possibility of removing some causal variants during LD-pruning). While not directly modeling LD is a limitation of this approach, the lower reliance on accurate LD information may be advantageous in some cases, for example when in-sample LD information is not available and only limited external reference data are present. On the other hand, we are also investigating methods to model LD in the SumVg framework. Since SumVg and LDSC are based on different modeling strategies and assumptions, and that the main focus of this study is the development of new SE/CI estimation approaches for SumVg (as well as applications to immune traits and presentation of a new R package), we shall leave carrying out a detailed comparison between SumVg and LDSC (or other SNP heritability estimation methods) for future work.

We have not investigated methods for SE estimation when raw genotype data are available. When raw data is available, one potential approach is to simply resample the individuals with a replacement (i.e., standard non-parametric bootstrap). However, such an approach is computationally intensive and its performance over methods based on summary statistics requires further research. The above resampling methods can also potentially be sped up by splitting the job into multiple processes to be run in parallel, although this approach has not been implemented in our software yet. We also wish to point out that, as the resampling methods were supposed to apply to GWAS summary data, in general the computational speed is fast, and the speed is not affected by sample sizes.

We have explored various approaches to construct CI, although we cannot yet find a single approach that yields an optimal CI with good coverage across all scenarios. We shall leave the development of more sophisticated and novel methodologies for CI construction for future works. For practical purposes, the union CI appears to perform well in terms of coverage across most scenarios (at the expense of wider CIs). On the other hand, we suspect that the issue of CI construction may not be unique to the SumVg approach; other methods for estimating SNP-based heritability typically require more stringent assumption on the distribution of effects, and/or that all SNPs contribute to heritability. The violation of such assumptions may lead to the estimates being biased and the inadequate coverage of CIs. Here we have proposed a bootstrap correction of bias, which indeed led to improvement in CI coverage in some cases, for example the standard CI under small sample sizes. Nevertheless, bootstrap correction showed a variable performance across different scenarios and did not always reduce bias in all cases. The above issues may warrant further studies.

Here we further highlight several important points to note and limitations of our framework. Regarding the SumVg estimator of total SNP-based heritability, one future research direction is to further explore its asymptotic theoretical properties. We did not pursue this direction here. Of note, the key difficulty in Equation (1) (i.e., the Tweedie's formula) is to estimate $f(x)$ and $f'(x)$ accurately. We primarily employed a kernel density estimator here, although other density estimation approaches may also be attempted. Notably, the kernel density estimator has been shown to be asymptotically consistent under certain assumptions [27]. In the paper by Efron [28], the asymptotic regret (Reg) of the empirical Bayes approach (i.e, using Tweedie's formula) was studied by comparing the Tweedie's estimate with the Bayes estimate of the true effect size, for a fixed value of z at z_0. It was shown that $\text{Reg}(z_0)$ tends towards zero as N tends towards infinity, and the regret depends on the squared error of $\widehat{l'}(z_0)$ as an estimator of $l'(z_0)$, where $l'(z) = \frac{d}{dz} \log f(z)$. Future theoretical studies of SumVg and other SNP heritability estimation methods are warranted.

In the current work, we assume that the summary statistics have been corrected for population stratification and other types of bias. If the original GWAS study suffered from bias, e.g., confounding, selection/ascertainment bias, sampling bias, bias due to missing data, etc., the resulting Vg estimate will also be affected. We suggest that the above bias should be carefully addressed at the design and/or analysis stage of the GWAS, for example by performing proper random sampling, inverse probability weighting to address selection bias [29], proper imputation of missing data, etc. As with any method, independent replication is also important.

Another limitation is that the proposed approach for calculating SNP heritability and SE/CI estimation may not work well for very small sample sizes. Since GWAS sample sizes are generally getting larger (most with $N > 5000$), we did not address the performance under very small sample sizes here. In such cases, both the SNP heritability and SE estimates may need to be viewed with greater caution. Meta-analysis of GWAS results across multiple studies may be recommended. Future work may also explore more innovative approaches to addressing small sample sizes, for example whether specifying a prior for the underlying effect sizes (δ) may help. (The current approach does not require any specification of the distribution of δ).

We also note that resampling methods often assume that the data points are independent of each other. In our study, prior to the analysis, we processed the data to remove strongly linked SNPs using LD pruning. The resulting SNPs are therefore roughly independent though some residual LD might remain. As a future direction, it may be useful to explore ways to fully tackle LD, for example by block bootstrap or jackknife [30]. However, external LD data from reference panels would be required, and there may be risks of LD mismatch between the studied and external samples. Further studies are required to investigate these issues.

Different resampling methods like bootstrap and jackknife may have different assumptions and applicability to different kinds of data. We have conducted relatively extensive simulations to compare performance of different methods across a range of heritability levels and sample sizes, which helps evaluate their applicability. We believe the proposed methods are generally applicable to most GWAS summary data. Note that the parametric bootstrap approaches assume that the observed data (z-statistics) are drawn from a certain specified parametric distribution. In our case, it is assumed that the δ and/or local fdr are estimated reasonably well. For small sample sizes, this assumption may not hold very well. The jackknife approaches do not require parametric assumptions; however, delete-one-jackknife has been shown to produce inconsistent variance estimators for non-smooth estimators such as the sample quantiles [31]. Delete-d-jackknife can resolve this problem, but the choice of d may not be straightforward. We suggest that multiple types of resampling methods should be performed; similar results across different methods may provide reassurance to the validity of results. Future work may include more extensive simulations for different genetic architectures and wider applications to complex traits.

There may be a concern that resampling methods may not handle extreme values or skewed distributions well. As discussed above, we recommend the GWAS should be conducted carefully in the first place. For example, skewed phenotypes may require transformation before analysis, and confounding or other kinds of bias need to be addressed. The SumVg method works on summary statistics. It is possible to perform further inverse-rank transformation to the summary statistics if the distribution is skewed or outliers are present, although this may create some bias to the Vg estimate. One may also trim the outlying z-statistics, and increasing the number of resamples may also help. The performance of these approaches will be a topic for further studies.

Importantly, we have also applied our approach to estimate the heritability of different cytokines, which play important roles in immune response and the pathogenesis of autoimmune, inflammatory and infectious diseases. Our analyses suggest that the studied cytokines are moderately heritable in general.

To summarize, SumVg is useful for triangulating evidence from different approaches to support conclusions regarding SNP-based heritability. We present novel methods of computing SE and CI and an easy-to-use software here, which we believe will be helpful for other researchers. Our application to the cytokine levels also sheds light on the genetic architecture of these clinically important immune traits.

4. Materials and Methods

4.1. Estimation of the Total Heritability Explained (Vg)

We previously proposed an approach [10] to estimate the sum of heritability explained by all variants on a GWAS panel. Our approach leverages Tweedie's formula for estimating the true underlying effect sizes of SNPs, based on the observed GWAS summary statistics. The principles are described in detail in the work by Efron [28].

4.1.1. Estimation of Total Vg Based on Tweedie's Formula

More specifically, assuming we have a large number of normally distributed variables (here *z*-statistics from a GWAS analysis), each with its own unobserved mean parameter δ_i, then

$$z_i \sim \mathcal{N}\left(\delta_i, \sigma^2\right) , \quad i = 1, 2, \dots, k$$

where *k* is the total number of variables. The attention is focused on the more extreme values, for example the top SNPs in high-dimensional genomics studies. As described by Efron [28], 'selection bias' may be at play here. Intuitively, the more extreme *z*-statistics might have been 'lucky' as random errors pushed them to deviate from zero; as such they can 'stand out' among the other *z*-statistics. In other words, the true underlying effect sizes of these top SNPs tend to be less extreme than the observed values. This phenomenon is also known as the 'winner's curse', for example see [32–34]. As a result, if we directly used the observed *z*-statistics to estimate the true effect sizes, the performance may not be optimal. Some form of 'correction' of the observed *z*-statistics are required.

Efron [28] proposed an empirical Bayes approach to reduce the selection bias, which was first described by Robbins [11] who attributed the ideas to Tweedie. The method assumes that

$$\delta \sim g(\cdot) \quad \text{and} \quad z|\delta \sim \mathcal{N}\left(\delta, \sigma^2\right)$$

In other words, we assume that δ was sampled from a prior 'density' $g(.)$, then $z \sim \mathcal{N}(\delta, \sigma^2)$ were observed, and the variance σ^2 was known. There are no assumptions on the form of the prior density *g*. According to the Tweedie's formula,

$$\mathrm{E}\{\delta \mid z\} = z + \sigma^2 l'(z), \text{where } l'(z) = \frac{d}{dz} log\, f(z)$$

In our setting of GWAS analyses, we assume $\sigma^2 = 1$, since we work with the summary *z*-statistics. We estimated the true or 'corrected' effect sizes of SNPs using

$$E\{\delta|z\} = z + \frac{f'(z)}{f(z)} \tag{1}$$

which is equivalent to the formula above when $\sigma^2 = 1$. Here *z* denotes the observed *z*-statistic, obtained from the estimated regression coefficient divided by the estimated SE (i.e., $\frac{\hat{\beta}}{\widehat{SE}}$). δ is the *z*-statistic derived by the *true* effect size divided by the estimated SE of the sample ($\beta_{\text{true}}/\widehat{SE}$), which can be considered a form of the 'standardized' true effect size. We previously proposed to employ a kernel density estimator to compute *f(z)* [10], which was shown to perform well in simulations. The total variance explained (Vg) can be obtained by converting the underlying effects δ to the Vg scale (see below and ref [10]).

4.1.2. Conversion of *z*-Statistics to Vg

For continuous traits, the conversion formula followed our previous work [10], which can be derived from ANOVA table of regression,

$$Vg = \frac{[E(\delta|z)]^2}{n - 2 + [E(\delta|z)]^2}$$

For binary outcomes, it is also possible to convert the *z*-statistics to Vg, provided that the estimated SE (or beta) and minor allele frequencies (MAF) of the SNPs, as well as the outcome prevalence, are available. We followed the methodology described in ref [35], which described how to convert coefficients from a logistic model to the liability scale. Note that the liability is assumed to have a variance of one. We followed Equation (4) from the above paper [35] to derive the coefficient (τ_1) under a liability scale. We converted τ_1 to the standardized coefficient ($\tau_{standard}$) by multiplying τ_1 by sqrt($2 \times$ MAF $\times$ (1 − MAF)), which is the standard deviation (SD) of the allelic count (coded as 0, 1, 2). Total variance explained is given by sum of the squared $\tau_{standard}$.

4.1.3. Assumptions

Regarding the assumptions of this approach, we emphasize that it does *not* require prior assumptions about the underlying distributions of the true effect sizes δ, which is an important advantage over other SNP-heritability estimation methods. On the other hand, we assume that the summary statistics have been corrected for population stratification or other confounding factors. The *z*-statistics are assumed to follow normal distributions; for very small samples sizes, rare variants, highly imbalanced case to control ratio, or highly skewed continuous outcomes, etc., caution should be taken as to whether the test statistic $\frac{\hat{\beta}}{\widehat{SE}}$ follows a normal distribution. We assume full GWAS summary statistics as input; if the summary statistics have been selected based on their significance levels (e.g., some GWAS only released the top SNPs, say top 10,000 SNPs), the proposed Tweedie's formula may not work well. The effect sizes may be overestimated in this case as the other SNPs have been selected for being significant.

4.1.4. An Alternative Conditional Estimator

We also proposed an alternative approach by evaluating the expected effect size conditioned on H1 (i.e., $\delta \neq 0$)

$$E(\delta|z, \; H_1) = E_1(\delta|z) = \frac{E(\delta|z)}{Pr(H_1|z)} = \frac{E(\delta|z)}{1 - fdr(z)} \tag{2}$$

where fdr is the local false discovery rate described in Efron [36]. The resulting estimate of Vg can be obtained by first converting $E(\delta|z, \; H_1)$ to the Vg scale (see Section 4.1.2), then multiply by $1 - fdr(z)$.

The conditional estimator, however, is prone to large random variations as it involves local fdr estimation of each SNP. In many subsequent applications of our heritability estimation method [15–17], the unconditional estimator (Equation (1)) was primarily employed. We shall hence focus on the unconditional estimator in this paper, although the resampling approaches described below can readily be applied to other estimators in our previous work [10] as well.

4.2. Estimation of the Standard Error (SE) of Vg

4.2.1. Standard and Delete-*d*-Jackknife to Estimate SE

In standard (delete-one) jackknife procedure [37], we estimate the standard error (SE) by leaving out one observation at a time. The SE is defined by

$$\widehat{SE}_{jack} = \sqrt{\frac{n-1}{n} \sum \left(\widehat{\theta}_{(i)} - \widehat{\theta}_{(.)} \right)^2}$$

where n is the sample size, $\widehat{\theta}_{(i)}$ is the parameter estimate from the sample with the ith observation removed and

$$\widehat{\theta}_{(.)} = \sum_{i=1}^{n} \frac{\widehat{\theta}_{(i)}}{n}$$

In our case, the parameter is the sum of heritability from all variants.

An extension is the delete-d-jackknife [31] where we leave out d observations at a time. There are in total $N = \binom{n}{d}$ possibilities of removing d out of n observations. In practice, N is usually very large. One may simply randomly repeat the procedure m times only ($m \leq N$) instead of exhausting all possibilities of removing d out of n observations. The standard error is given by

$$\widehat{SE}_{del-d-jack} = \sqrt{\frac{n-d}{dm} \sum_{v=1}^{m} \left(\widehat{\theta}_{S_v} - \frac{1}{m} \sum_{v=1}^{m} \widehat{\theta}_{S_v} \right)^2}$$

where $\widehat{\theta}_{S_v}$ denotes the parameter estimate in the vth jackknife replicate where d observations are left out. The delete-d-jackknife (when $d > 1$) works better than the standard jackknife for non-smooth parameters like the median [31].

There are no clear rules on the choice of d in delete-d-bootstrap. Chatterjee [38] suggested $n/5$ as a reasonable choice for d based on the consideration of efficiency and likely model conditions. We followed the suggestion by Chatterjee [38] and set d as $n/5$ (=20,000) in all simulations.

4.2.2. Parametric Bootstrap Approaches for Estimating SE

In parametric bootstrap, in each replication we simulated z-statistics based on $\widehat{\delta}$, the 'corrected' z-statistics from original sample (this method is referred to as 'paraboot'). We have

$$z_{i,b} \sim N\left(\widehat{\delta}_i, 1 \right)$$

where $z_{i,b}$ denotes the ith z-statistic in the bth bootstrap replicate. For small effects, the $\widehat{\delta}$ will be shrunken towards zero.

We further proposed a modified approach by also considering the local fdr (i.e., probably of null given z) of each z-statistic. In each replicate, we simulated z-statistics according to the following scheme:

$$z_{i,b} \sim N\left(\widehat{z}_i, 1 \right) \text{ with a probability of } 1 - \widehat{fdr}(z_i)$$

$$z_{i,b} \sim N(0, 1) \text{ with a probability of } \widehat{fdr}(z_i)$$

where $\widehat{z}_i$ denotes the observed z-statistics. The standard error is then computed from the simulated z-statistics. This method is referred to as "fdrboot1".

Alternatively, one may employ the corrected z-statistics instead of the observed z-statistics as the mean in each simulation, i.e.,

$$z_{i,b} \sim N\left(\widehat{\delta}_i, 1 \right) \text{ with a probability of } 1 - \widehat{fdr}(z_i)$$

$$z_{i,b} \sim N(0, 1) \text{ with a probability of } \widehat{fdr}(z_i)$$

The method is also referred to as 'fdrboot2'.

4.3. Construction of Confidence Intervals (CIs): An Exploratory Analysis

The construction of a proper CI is a more demanding task as it requires the unbiasedness of the estimate and correct estimation of the variability of the estimate. Given the difficulty of constructing accurate CIs, here we consider CI estimation as a secondary or exploratory analysis which requires further investigation and methodological development. We have explored a few approaches as described below.

4.3.1. Normal Approximation (Standard Approach)

Firstly, we explored the standard approach for constructing the 95% CI by using normal approximation, i.e., $\widehat{Vg} \pm z_{0.975} \times \widehat{SE}(Vg)$, where $z_{0.975}$ is the quantile of a standard normal distribution at the 97.5th percentile. Assuming a polygenic model, the total heritability is the sum of variance explained contributed by many variants of small to modest effect sizes. Hence, it is reasonable to assume normality according to the central limit theorem (as is assumed by other SNP-heritability estimation tools). We examined the performance of different CIs, with SE determined by various methods. Empirically, we found that SE computed by the delete-d-jackknife performed reasonably well.

On the other hand, we also explored this using bootstrap to correct for bias of the point estimates of Vg. In brief, the bias can be estimated by [39]

$$\widehat{Bias} = \overline{\theta^*} - \widehat{\theta}$$

where $\widehat{\theta}$ denotes the observed Vg, and $\overline{\theta^*}$ is the mean of the bootstrapped estimates of Vg. The bias-corrected estimator of Vg is given by

$$Bias\ corrected\ \theta = \widehat{\theta} - \widehat{Bias}$$

The 95% CI is then based on $Bias\ corrected\ \widehat{Vg} \pm z_{0.975} \times \widehat{SE}$. Since we proposed 3 bootstrap procedures, there were 3 bootstrap bias-corrected CIs based on normal approximation. The standard CI without bias correction was also included as another estimator.

4.3.2. Percentile Approach

Secondly, we explored the percentile CI approach, namely construction of 95% CIs based on the 2.5th and 97.5th percentiles of the bootstrapped Vg. Again, bias correction can be applied as follows

$$\text{Lower 95% CI} = 2\widehat{\theta} - \theta^*_{(0.975)}$$

$$\text{Upper 95% CI} = 2\widehat{\theta} - \theta^*_{(0.025)}$$

where $\theta^*_{(0.025)}$ and $\theta^*_{(0.975)}$ are the 2.5th and 97.5th percentiles of the bootstrapped replicates of Vg, respectively. Bias correction was based on the same bootstrap method that was used to derive the percentiles. Again, we also included the percentile CIs without bias correction.

4.3.3. Union CI

Thirdly, we explored a more robust CI estimator by taking the union of individual CIs (UCI). The union of multiple CIs is constructed by taking the minimum of the lower CIs across different methods as the final lower CI, and the maximum of different upper CIs as the final upper CI. This union approach can ensure better robustness if CI construction approaches perform differently under different scenarios. The UCI method has been widely employed in instrumental variables regression to improve robustness of results in the presence of pleiotropy [40].

In summary, the following methods were explored:

1.   Normal approximation (standard approach), without bias correction (one estimator) or with bootstrap bias correction (3 estimators), then take the union of CIs;

2. Percentile approach, without bias correction (3 estimators) and with bias correction (3 estimators), then take the union of CIs;
3. Union of the final CI obtained from 1 and 2.

4.4. Simulation Studies

We compare the SE estimated from the above methods with the 'true' SE obtained from one hundred simulations with known data generating distributions. The details of the simulations is as follows [10]. Briefly, a gamma distribution was used to simulate three levels of variance explained (Vg = 0.101, 0.191, 0.295), which were converted to true effect sizes (δ). Z-statistics for 100,000 independent SNPs (0.5% were non-null) with different sample sizes (N = 5000, 10,000, 20,000, 50,000, 100,000, 200,000) were then simulated as input for SumVg following the distribution $N(\delta,1)$. Two hundred replicates were run for each bootstrap or jackknife procedure. We focus on quantitative traits in our simulations, but the results should most likely apply to binary traits as well, as the only difference in these two scenarios is the formula to convert z to variance explained (Vg). The performance of different methods for CI construction was also evaluated.

4.5. Application to Immune Traits

A selected set of immune-related traits (levels of cytokines/growth factors) were included for study, based on the GWAS by Ahola-Olli et al. [24]. We selected 12 continuous immune traits with (1) sample size $N > 5000$ and (2) very low ($\leq$3%) or negative SNP-based heritability estimated by LDSC. The LDSC heritability were based on pre-calculated values from GWASAtlas (https://atlas.ctglab.nl/; accessed 1 May 2023). SNPs in strong LD were removed using the PLINK command "--indep-pairwise 100 25 r2" with a series of r^2 thresholds (0.1, 0.05, 0.025, 0.01, 0.005, 0.002, 0.001). The 1000G Phase3 EUR sample was used as the reference panel to calculate LD among variants. Independent SNPs with MAF > 0.01 were then applied to SumVg.

References

1. Yang, J.; Benyamin, B.; McEvoy, B.P.; Gordon, S.; Henders, A.K.; Nyholt, D.R.; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **2010**, *42*, 565–569. [CrossRef] [PubMed]
2. Speed, D.; Hemani, G.; Johnson, M.R.; Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **2012**, *91*, 1011–1021. [CrossRef] [PubMed]
3. Bulik-Sullivan, B.K.; Loh, P.R.; Finucane, H.K.; Ripke, S.; Yang, J.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N.; Daly, M.J.; Price, A.L.; Neale, B.M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **2015**, *47*, 291–295. [CrossRef] [PubMed]
4. Speed, D.; Balding, D.J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **2019**, *51*, 277–284. [CrossRef]
5. Zhu, H.; Zhou, X. Statistical methods for SNP heritability estimation and partition: A review. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1557–1568. [CrossRef]
6. Barry, C.-J.S.; Walker, V.M.; Cheesman, R.; Davey Smith, G.; Morris, T.T.; Davies, N.M. How to estimate heritability: A guide for genetic epidemiologists. *Int. J. Epidemiol.* **2023**, *52*, 624–632. [CrossRef]
7. Zuk, O.; Hechter, E.; Sunyaev, S.R.; Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1193–1198. [CrossRef]
8. Brandes, N.; Weissbrod, O.; Linial, M. Open problems in human trait genetics. *Genome Biol.* **2022**, *23*, 131. [CrossRef] [PubMed]
9. Young, A.I. Solving the missing heritability problem. *PLoS Genet.* **2019**, *15*, e1008222. [CrossRef]
10. So, H.C.; Li, M.; Sham, P.C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.* **2011**, *35*, 447–456. [CrossRef]
11. Robbins, H. An empirical Bayes approach to statistics. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Cambridge, UK, 26–31 December 1954, July and August 1955; University of California Press: Berkeley, CA, USA; Los Angeles, CA, USA, 1956; Volume 1, pp. 157–163.
12. Brown, L.D. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **1971**, *42*, 855–903. [CrossRef]
13. Efron, B. Empirical Bayes estimates for large-scale prediction problems. *J. Am. Stat. Assoc.* **2009**, *104*, 1015–1028. [CrossRef]
14. Zhang, Y.; Cheng, Y.; Jiang, W.; Ye, Y.; Lu, Q.; Zhao, H. Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Brief. Bioinform.* **2021**, *22*, bbaa442. [CrossRef] [PubMed]
15. Benke, K.S.; Nivard, M.G.; Velders, F.P.; Walters, R.K.; Pappa, I.; Scheet, P.A.; Xiao, X.; Ehli, E.A.; Palmer, L.J.; Whitehouse, A.J.; et al. A genome-wide association meta-analysis of preschool internalizing problems. *J. Am. Acad. Child. Adolesc. Psychiatry* **2014**, *53*, 667–676.e667. [CrossRef] [PubMed]
16. Lubke, G.H.; Hottenga, J.J.; Walters, R.; Laurin, C.; de Geus, E.J.; Willemsen, G.; Smit, J.H.; Middeldorp, C.M.; Penninx, B.W.; Vink, J.M.; et al. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol. Psychiatry* **2012**, *72*, 707–709. [CrossRef]
17. van Beek, J.H.; Lubke, G.H.; de Moor, M.H.; Willemsen, G.; de Geus, E.J.; Hottenga, J.J.; Walters, R.K.; Smit, J.H.; Penninx, B.W.; Boomsma, D.I. Heritability of liver enzyme levels estimated from genome-wide SNP data. *Eur. J. Hum. Genet.* **2014**, *23*, 1223–1228. [CrossRef]
18. Hibar, D.P.; Stein, J.L.; Renteria, M.E.; Arias-Vasquez, A.; Desrivieres, S.; Jahanshad, N.; Toro, R.; Wittfeld, K.; Abramovic, L.; Andersson, M.; et al. Common genetic variants influence human subcortical brain structures. *Nature* **2015**, *520*, 224–229. [CrossRef]
19. Paternoster, L.; Standl, M.; Waage, J.; Baurecht, H.; Hotze, M.; Strachan, D.P.; Curtin, J.A.; Bonnelykke, K.; Tian, C.; Takahashi, A.; et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **2015**, *47*, 1449–1456. [CrossRef]
20. Lo, M.T.; Hinds, D.A.; Tung, J.Y.; Franz, C.; Fan, C.C.; Wang, Y.; Smeland, O.B.; Schork, A.; Holland, D.; Kauppi, K.; et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **2017**, *49*, 152–156. [CrossRef] [PubMed]
21. Minica, C.C.; Verweij, K.J.H.; van der Most, P.J.; Mbarek, H.; Bernard, M.; van Eijk, K.R.; Lind, P.A.; Liu, M.Z.; Maciejewski, D.F.; Palviainen, T.; et al. Genome-wide association meta-analysis of age at first cannabis use. *Addiction* **2018**, *113*, 2073–2086. [CrossRef] [PubMed]
22. Ahluwalia, T.S.; Prins, B.P.; Abdollahi, M.; Armstrong, N.J.; Aslibekyan, S.; Bain, L.; Jefferis, B.; Baumert, J.; Beekman, M.; Ben-Shlomo, Y.; et al. Genome-wide association study of circulating interleukin 6 levels identifies novel loci. *Hum. Mol. Genet.* **2021**, *30*, 393–409. [CrossRef]
23. Shin, S.H.; Park, S.; Wright, C.; D'Astous, V.A.; Kim, G. The Role of Polygenic Score and Cognitive Activity in Cognitive Functioning Among Older Adults. *Gerontologist* **2021**, *61*, 319–329. [CrossRef] [PubMed]
24. Ahola-Olli, A.V.; Würtz, P.; Havulinna, A.S.; Aalto, K.; Pitkänen, N.; Lehtimäki, T.; Kähönen, M.; Lyytikäinen, L.P.; Raitoharju, E.; Seppälä, I.; et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am. J. Hum. Genet.* **2017**, *100*, 40–50. [CrossRef] [PubMed]

25. Turner, M.D.; Nedjai, B.; Hurst, T.; Pennington, D.J. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochim. Biophys. Acta (BBA)—Mol. Cell Res.* **2014**, *1843*, 2563–2582. [CrossRef] [PubMed]

26. Steinsaltz, D.; Dahl, A.; Wachter, K.W. On Negative Heritability and Negative Estimates of Heritability. *Genetics* **2020**, *215*, 343–357. [CrossRef] [PubMed]

27. Wied, D.; Weißbach, R. Consistency of the kernel density estimator: A survey. *Stat. Pap.* **2012**, *53*, 1–21. [CrossRef]

28. Efron, B. Tweedie's formula and selection bias. *J. Am. Stat. Assoc.* **2011**, *106*, 1602. [CrossRef] [PubMed]

29. Carry, P.M.; Vanderlinden, L.A.; Dong, F.; Buckner, T.; Litkowski, E.; Vigers, T.; Norris, J.M.; Kechris, K. Inverse probability weighting is an effective method to address selection bias during the analysis of high dimensional data. *Genet. Epidemiol.* **2021**, *45*, 593–603. [CrossRef]

30. Horowitz, J.L. Bootstrap methods in econometrics. *Annu. Rev. Econ.* **2019**, *11*, 193–224. [CrossRef]

31. Shao, J.; Wu, C.J. A general theory for jackknife variance estimation. *Ann. Stat.* **1989**, *17*, 1176–1197. [CrossRef]

32. Zhong, H.; Prentice, R.L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **2008**, *9*, 621–634. [CrossRef] [PubMed]

33. Sun, L.; Bull, S.B. Reduction of selection bias in genomewide studies by resampling. *Genet. Epidemiol. Off. Publ. Int. Genet. Epidemiol. Soc.* **2005**, *28*, 352–367. [CrossRef] [PubMed]

34. Zöllner, S.; Pritchard, J.K. Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **2007**, *80*, 605–615. [CrossRef] [PubMed]

35. Gillett, A.C.; Vassos, E.; Lewis, C.M. Transforming summary statistics from logistic regression to the liability scale: Application to genetic and environmental risk scores. *Hum. Hered.* **2019**, *83*, 210–224. [CrossRef]

36. Efron, B.; Tibshirani, R.; Storey, J.D.; Tusher, V. Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.* **2001**, *96*, 1151–1160. [CrossRef]

37. Miller, R.G. The jackknife—A review. *Biometrika* **1974**, *61*, 1–15.

38. Chatterjee, S. Another look at the jackknife: Further examples of generalized bootstrap. *Stat. Probab. Lett.* **1998**, *40*, 307–319. [CrossRef]

39. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.

40. Conley, T.G.; Hansen, C.B.; Rossi, P.E. Plausibly exogenous. *Rev. Econ. Stat.* **2012**, *94*, 260–272. [CrossRef]