Accounting for uncertainty in residual variances improves calibration of the "Sum of Single Effects" model for small sample sizes

William R.P. Denault[1], Peter Carbonetto[1], Ruixi Li[2], The Alzheimer's Disease Functional Genomics Consortium, Gao Wang[2,3], and Matthew Stephens[1,4]

[1]Department of Human Genetics, University of Chicago, Chicago, IL, USA
[2]Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
[3]Department of Neurology, Columbia University, New York, NY, USA
[4]Department of Statistics, University of Chicago, Chicago, IL, USA
[*]Correspondence: William R.P. Denault, wdenault@uchicago.edu; Matthew Stephens, mstephens@uchicago.edu

Abstract

The Sum of Single Effects (SuSiE) model, and a fitting procedure based on variational approximation, were recently introduced as a way to capture uncertainty in variable selection in multiple linear regression. This approach is particularly well suited to cases where variables are highly correlated, and it has been rapidly adopted for genetic fine-mapping studies. Here we show that, in small studies (<50 samples), the original SuSiE fitting procedure can produce substantially higher rates of false positive findings than it does in larger studies. We show that a simple alternative fitting procedure, which takes account of uncertainty in the residual variance, notably improves performance in small sample studies.

1. Introduction

Wang et al [1] introduced the "Sum of Single Effects" (SuSiE) model as a way to capture uncertainty in variable selection in multiple linear regression. This work was motivated by settings like genetic fine-mapping which often involves very highly correlated variables, and where it is difficult to confidently identify the "relevant" variables (those with a non-zero regression coefficient). Wang et al pointed out that, even in such settings, it may be possible, and useful, to narrow down the potentially relevant variables. They suggest doing this by identifying sets of (highly correlated) variables that, with high probability, contain at least one variable with a non-zero coefficient. [1] refer to these sets of variables as "Credible Sets" (CSs), and the ability of SuSiE to efficiently generate reliable CSs is a key feature of the method. Since its introduction, several papers have extended the method in various ways [2–6].

Here we highlight a limitation of the original SuSiE method: it can produce badly miscalibrated CSs when applied to data sets with small sample sizes ($n \leq 50$). We show that this miscalibration occurs even in the simplest case where exactly one variable is assumed to have a non-zero coefficient. While many fine-mapping studies are large, this miscalibration issue may nonetheless be important in smaller studies such as pilot experiments, complex experimental designs (e.g., dynamic eQTLs from [7], $n = 19$, and more recently [8], $n = 53$), costly multi-modal -omics experiments (e.g., multi-modal QTL mapping in influenza-infected cells [9], $n = 35$), and studies of rare cell types (e.g., eQTL mapping in microglia from the subventricular zone [10], $n = 47$).

We show that this miscalibration can be greatly improved by a simple modification of the procedures in [1] to account for uncertainty in the residual variance parameter. Specifically, whereas the original SuSiE method estimates the residual variance by maximum-likelihood, here we introduce methods to integrate it out under a diffuse prior, an idea that has its roots in the classical Bayesian literature [11–14]. This simple modification substantially improves CS calibration in simulations, and provides more realistic results when applying the SuSiE model to small-sample settings (Fig. 1).

2. Single effect regression accounting for uncertainty in residual variance estimates

The Sum of Single Effects (SuSiE) model is built on the simpler "Single Effect Regression" (SER) model, which assumes that *exactly one* of the variables in the regression has a non-zero effect on the outcome. The SER model from [1] is

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \epsilon \\ \boldsymbol{\beta} &= \boldsymbol{\gamma}\beta \\ \boldsymbol{\gamma} &= \text{Multinom}(1, \boldsymbol{\pi}) \\ \epsilon &\sim N(0, \sigma^2) \\ \beta &\sim N(0, \sigma_0^2\sigma^2), \end{aligned} \tag{1}$$

where $\boldsymbol{y}$ is the vector of observed outcomes, $\boldsymbol{X}$ is a matrix of explanatory variables, $\boldsymbol{\beta}$ is the vector containing the effects of these variables on the phenotype, and ϵ is Gaussian noise with variance σ^2. Note that $\boldsymbol{\gamma}$ is a vector in which exactly one entry is different from zero. The prior probability for which entry is non-zero is encoded by the vector $\boldsymbol{\pi}$ (it is a vector with non-negative entries that sum to one). The hyperparameter σ_0^2 determines the prior on the effect size relative to the residual variance. In [1], the residual variance σ^2 was estimated by maximizing the likelihood (or, in SuSiE, by maximizing the "evidence lower bound"). Given σ^2 and σ_0^2, it is straightforward to compute the posterior distribution on $\boldsymbol{\beta}$ analytically, essentially by summing over all the possible values for $\boldsymbol{\gamma}$ [1, 12].

Here we augment the SER model by placing a (conjugate) inverse-gamma prior on the residual variance:

$$\sigma^2 \sim \text{IG}\left(\tfrac{\kappa}{2}, \tfrac{\lambda}{2}\right). \tag{2}$$

This augmented SER is, in essence, the model considered in [12], and so we call it the "SS-SER" ("SS" being short for "Servin and Stephens"). Given the hyperparameters κ, λ and σ_0^2, it is straightforward to compute the posterior distribution on $\boldsymbol{\beta}$ analytically. Servin and Stephens [12] showed that this posterior distribution has a sensible limit as $\kappa \to 0, \lambda \to 0$, and we use that limiting posterior distribution here.

2.1. Extension to the SuSiE model: SuSiE with the SS-SER model

The SER model is limited because it allows for only a single non-zero effect. Therefore, Wang et al [1] suggested allowing for $L \geq 1$ effects by summing L SER models, yielding the "sum of single effects" (SuSiE) model. They developed a fitting procedure, "iterative Bayesian stepwise selection" (IBSS), which fits the SuSiE model by *iteratively* fitting a sequence of SER models. They showed that this IBSS algorithm, when used with their SER, corresponds to optimizing a variational approximation [15] to the posterior distribution. In brief, here we use the same IBSS algorithm as in [1], but replace their SER (which fixes σ at a point estimate), with the SS-SER (which integrates out σ); see the Supporting derivations for details. Unlike the original IBSS algorithm, this modified IBSS is not guaranteed to be exactly optimizing a variational approximation to the posterior distribution; however, our empirical results below show that in small-sample studies it can produce better CS calibration while maintaining the simple computations of the original IBSS algorithm.

3. Results

3.1. Numerical simulations

We compared our new approach (SuSiE with the *SS-SER*) with the original approach (SuSiE with the *default SER*) on simulated data. For each simulation, we sampled genotypes by randomly selecting loci from the 1000 Genomes Project Phase 3 [16], including only unrelated individuals. Within the selected loci, we retained biallelic single nucleotide polymorphisms (SNPs) with minor allele frequencies $\geq$5%. From the remaining SNPs, we randomly selected 1–10 causal variants that collectively explained a specified proportion of the phenotypic variance ($h^2 = 25\%, 30\%, 50\%, 75\%$). These steps were repeated using different sample sizes ($N = 20, 30, 50, 75, 100$). We ran both methods using an upper bound of

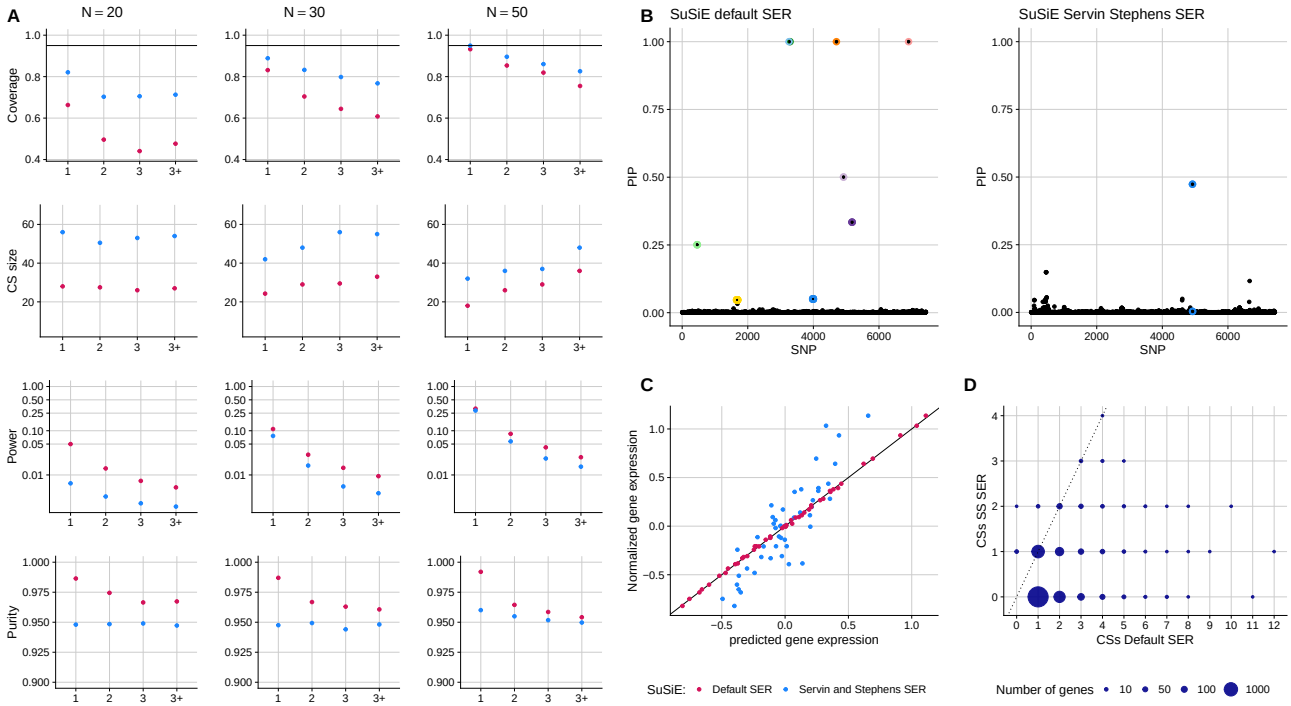

Fig. 1. Panel A: Comparison of 95% CSs from SuSiE with the default SER (red) and SuSiE with the SS-SER (blue) in a simulation setting where the genotype accounts for 25% of the trait variance. The top row displays the observed empirical coverage of the two methods for *N*= 20,30, and 50. The black line represents the target 95% coverage. Coverage was calculated as $1 - \frac{FP}{TP+FP}$ where FP and TP denote the number of false positives and true positives, respectively. The second row gives the mean CS size. The third row gives the power. Note the y-axis is on a logarithmic scale. The fourth row presents the median CS purity. **Panels B, C:** Illustration of severe miscalibration of SuSiE with the default SER in fine-mapping *ADAM17* expression using data from microglia in the subventricular zone from [10], *N* = 47. **Panel B** shows the estimated posterior inclusion probabilities (PIP) and credible sets (CS) for both SuSiE with the default SER and SS-SER. The x-axis represents the SNP index, while the y-axis gives the PIPs. Colors correspond to different CSs. SuSiE with the default SER identifies 9 CSs, whereas SuSiE with SS-SER reports 1 CS (this CS overlaps a CS from SuSiE with default SER). **Panel C** compares the predicted *ADAM17* gene expression under both models against normalized gene expression. The default SER exhibits explains $> 99\%$ of the variance, suggesting overfitting, while the SS-SER model provides a more realistic fit (68% variance explained). **Panel D:** summary of genome-wide fine-mapping using RNA-seq data from microglia in the subventricular zone [10]. It compares of the number of 95% credible sets (CSs) reported by SuSiE using the default SER (x-axis) versus SuSiE with the SS-SER (y-axis). For improved visualization, the point at (0,0) is removed; in 12,232 genes, both methods give no CSs.

$L = 10$ causal variants, and with the purity filtering step from [1], which removed "low purity" CSs (CSs that contain pairs of variables that have correlation less than 0.5). See Methods for details.

Our results (Fig. 1A, Fig. 2) show that, for small sample size studies ($n \leq 50$), SuSiE with the default SER had consistently poor CS coverage, substantially below the nominal 95% level. By comparison, using the SS-SER substantially improves coverage (although not attaining 95%). For example, with $N = 20$ and $h^2 = 0.25$, the default SER CS coverage can be as low as 44.1%, increasing to 82.1% using the SS-SER (Fig. 1A). The reason for this improved coverage is that SS-SER tends to produce larger CSs (Fig. 1A, Fig. 3) with lower purity (Fig. 1A, Fig. 4). Consequently more CSs are discarded by the purity filter, and so it also displays a corresponding drop in "power" (Fig. 1A, Fig. 5). It seems likely that this loss of power and localization are an inevitable consequence of improving the calibration.

To gauge the impact of estimating the number of causal variants (as well as the purity filtering step), we conducted an additional set of simpler simulations where a single variant affects the trait, and assessed CS coverage when (i) allowing SuSiE to estimate the number of causal effects (setting

$L = 10$ as above), and (ii) running SuSiE with the correct number of causal effects ($L = 1$). We ran (i) and (ii) both with and without purity filtering. Note that fitting SuSiE with $L = 1$ does not involve any variational approximation, so (ii) removes any potential error in this approximation. We summarize the results of these simulations in Figures 6 and 7.

In these simpler simulations, the mis-calibration issue appears only with the smallest sample sizes ($n = 20, 30$) and smallest effect size $h^2 = 0.25, 0.3$, and only in the cases with the purity filtering. Having SuSiE estimate the number of causal effects has only a minor impact on the calibration in all cases. These results may seem to suggest avoiding the purity filter altogether; however, in practice this leads to very large numbers of "noninformative CSs" that contain very large numbers of variables. These noninformative CSs technically improve calibration because, by including almost every variable, they usually include a relevant variable by chance. However, outputting large numbers of uninformative CSs to improve calibration in this way is typically undesirable. Our recommendation for small studies is therefore to maintain the purity filter, use the SS-SER, and bear in mind that the resulting CSs may still be somewhat less reliable than for large studies.

3.2. Fine mapping Aracena et al. and Lopes et al. data

To compare results from the different methods on real data, we fine-mapped normalized gene expression (RNA-seq) data from two studies. The first dataset from Aracena et al [9], was used to examine the transcriptional response in primary macrophages (n = 35) before and after infection with influenza. The second dataset from Lopes et al [10] was used to investigate expression QTLs in microglia, a key cell type involved in brain aging.

For each gene, we selected the SNPs within ±200 kb of the gene body with MAF > 0.05, and fine-mapped these SNPs using both approaches (SuSiE with the default SER, and SuSiE with the SS-SER). For the Aracena et al [9] study, across 13,680 fine-mapped genes, default SuSiE reported 1,020 unique CSs, whereas SuSiE with SS-SER reported 389 CSs. Among the genes where SuSiE with SS-SER identified a single CS (381 out of 389), this CS always overlapped with a CS reported by SuSiE with the default SER.

For the Lopes et al [10] study, across 17,132 fine-mapped genes, default SuSiE identified 6,918 CSs, while SuSiE with SS-SER reported 1,556 CSs. Again, nearly all genes with a single CS reported by SuSiE with SS-SER (844 out of 858) had overlapping CSs from SuSiE with the default SER. In the fine-mapping analysis of [10], SuSiE exhibited clear signs of miscalibration (Fig. 1B–D), with 146 genes having 4 or more CSs, including up to 12 CSs for *GAL3ST2*. By contrast, SuSiE with the SS-SER reported only one gene with 4 CSs.

4. Conclusion

We have highlighted a limitation of the SuSiE method: it often provides poorly calibrated CSs in studies with sample sizes. We also provided a simple modification that uses the SS-SER to better account for uncertainty in the residual variance parameter, and improves CS calibration. Our method retains the computational advantages of SuSiE and can be applied to summary statistics [17], making it practical for a wide range of fine-mapping studies. While our approach improves credible set coverage, some miscalibration remains in some settings, and further improving performance for small-sample studies could be an area for future work.

References

[1] Wang, G., A. Sarkar, P. Carbonetto, and M. Stephens (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society, Series B 82*(5), 1273–1300.

[2] Gao, B. and X. Zhou (2024). MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nature Genetics 56*, 170–179.

[3] Rossen, J., H. Shi, B. J. Strober, M. J. Zhang, M. Kanai, Z. R. McCaw, L. Liang, O. Weissbrod, and A. L. Price (2024). MultiSuSiE improves multi-ancestry fine-mapping in All of Us whole-genome sequencing data. *medRxiv*, doi:10.1101/2024.05.13.24307291.

[4] Yuan, K., R. J. Longchamps, A. F. Pardiñas, M. Yu, T.-T. Chen, S.-C. Lin, et al. (2024). Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nature Genetics 56*(9), 1841–1850.

[5] Zou, Y., P. Carbonetto, D. Xie, G. Wang, and M. Stephens (2023). Fast and flexible joint fine-mapping of multiple traits via the Sum of Single Effects model. *bioRxiv*, doi:10.1101/2023.04.14.536893.

[6] Cui, R., R. A. Elzur, M. Kanai, J. C. Ulirsch, O. Weissbrod, M. J. Daly, B. M. Neale, Z. Fan, and H. K. Finucane (2024). Improving fine-mapping by modeling infinitesimal effects. *Nature Genetics 56*, 162–169.

[7] Strober, B., R. Elorbany, K. Rhodes, N. Krishnan, K. Tayeb, A. Battle, and Y. Gilad (2019). Dynamic genetic regulation of gene expression during cellular differentiation. *Science 364*(6447), 1287–1290.

[8] Popp, J. M., K. Rhodes, R. Jangi, M. Li, K. Barr, K. Tayeb, A. Battle, and Y. Gilad (2024). Cell type and dynamic state govern genetic regulation of gene expression in heterogeneous differentiating cultures. *Cell Genomics 4*(12), 100701.

[9] Aracena, K. A., Y.-L. Lin, K. Luo, A. Pacis, S. Gona, Z. Mu, et al. (2024). Epigenetic variation impacts individual differences in the transcriptional response to influenza infection. *Nature Genetics 56*(3), 408–419.

[10] Lopes, K. P., G. J. L. Snijders, J. Humphrey, A. Allan, M. A. M. Sneeboer, E. Navarro, B. M. Schilder, R. A. Vialle, M. Parks, R. Missall, W. van Zuiden, F. A. J. Gigase, R. Kübler, A. B. van Berlekom, E. M. Hicks, C. Böttcher, J. Priller, R. S. Kahn, L. D. de Witte, and T. Raj (2022, Jan). Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nature Genetics 54*, 4–17.

[11] Gönen, M., W. O. Johnson, Y. Lu, and P. H. Westfall (2005). The Bayesian two-sample t test. *American Statistician 59*(3), 252–257.

[12] Servin, B. and M. Stephens (2007). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genetics 3*(7), e114.

[13] Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*(481), 410–423.

[14] Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review 16*, 225–237.

[15] Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: a review for statisticians. *Journal of the American Statistical Association 112*(518), 859–877.

[16] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature 526*(7571), 68–74.

[17] Zou, Y., P. Carbonetto, G. Wang, and M. Stephens (2022). Fine-mapping from summary data with the "Sum of Single Effects" model. *PLoS Genetics 18*(7), e1010299.

Additional figures

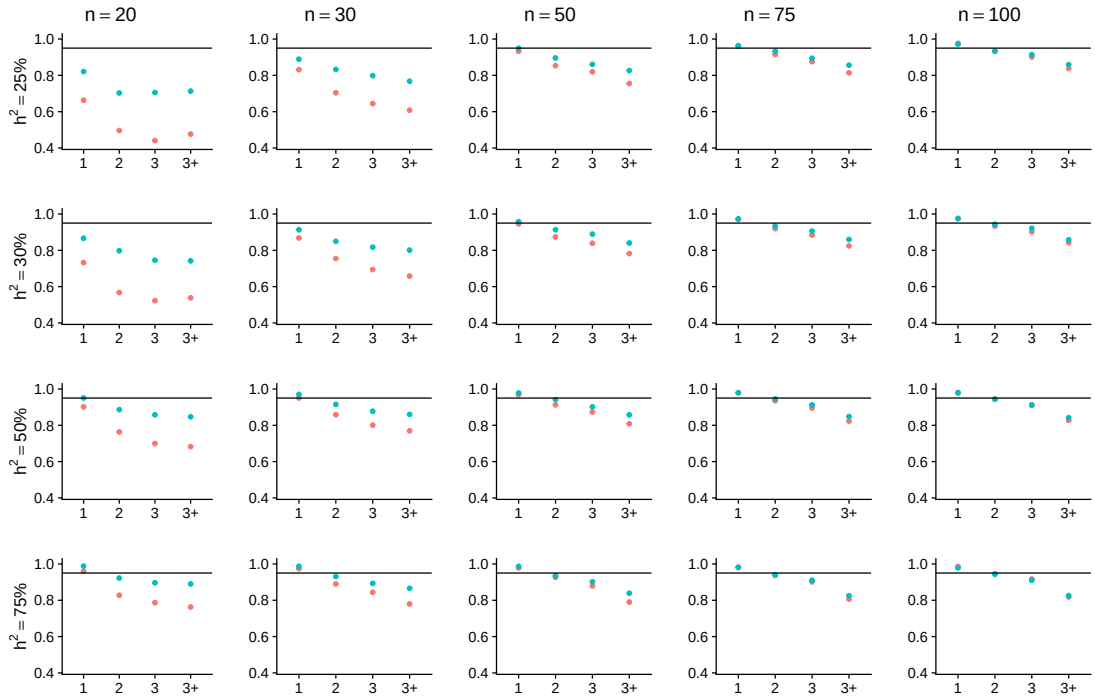

Fig. 2. Comparison of coverage (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue). The black line represents the target 95% coverage. Coverage was calculated as $1 - \frac{FP}{TP+FP}$ where FP and TP denote, respectively, the number of false positives and true positives.

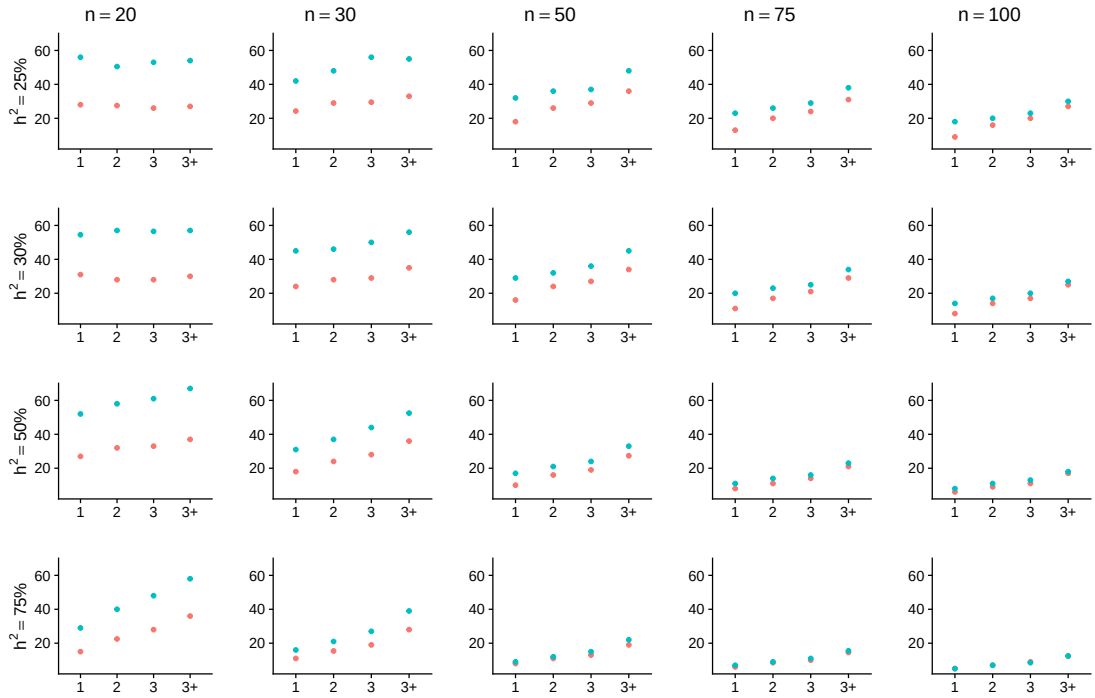

Fig. 3. Comparison of mean CS size (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue).

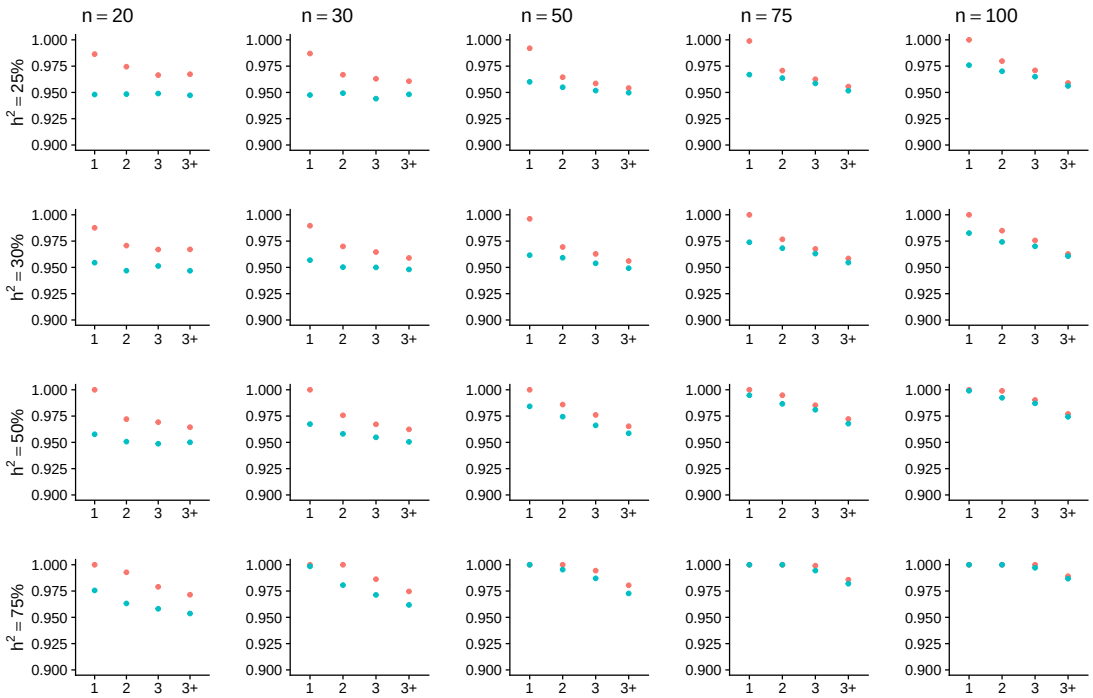

Fig. 4. Comparison of median CS purity (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue).

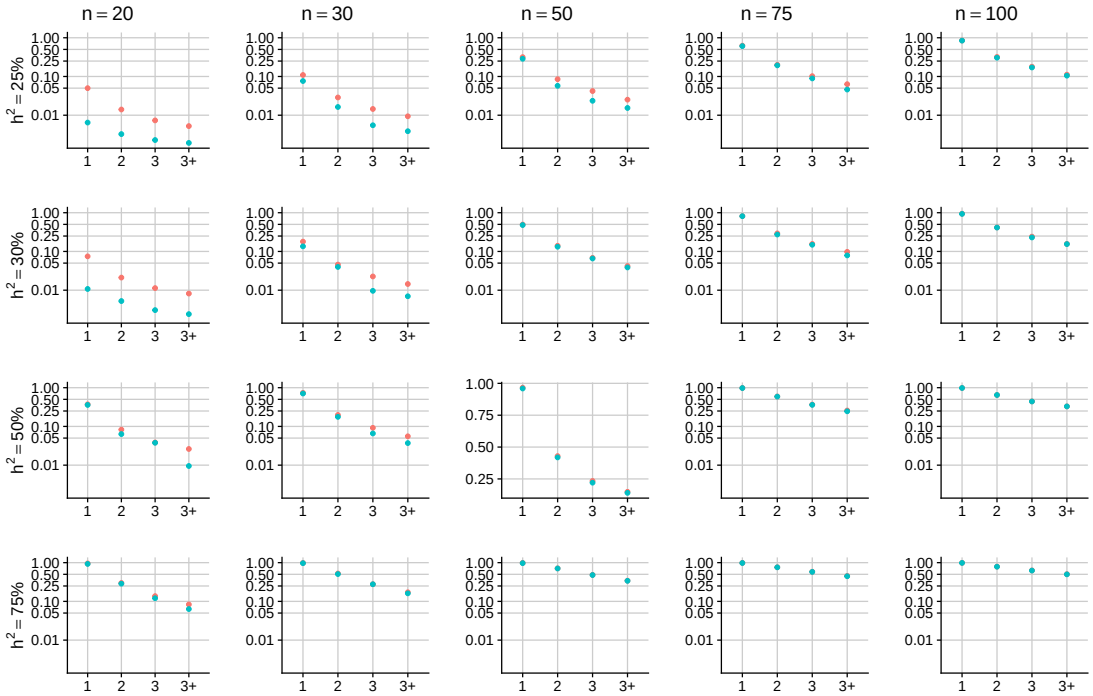

Fig. 5. Comparison of power (y-axis) for SuSiE using the default SER (red) and the SS-SER (blue). Note the y-axis is on a logarithmic scale.

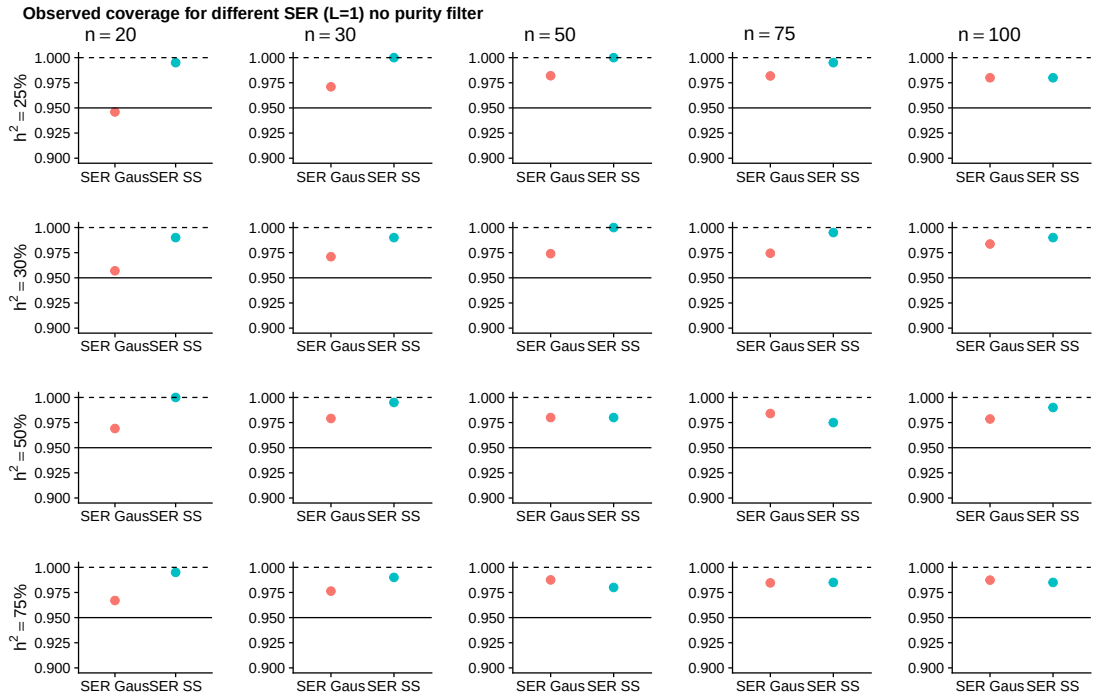

Fig. 6. Comparison of coverage between SuSiE using the default SER (red) and SuSiE using the SS-SER (blue) under a single causal SNP scenario without purity filtering. The solid line denotes the target 95% coverage level, while the dashed line indicates 100% coverage. Simulations were conducted with the $L = 1$, a configuration under which SuSiE computes exact posteriors for both SERs.

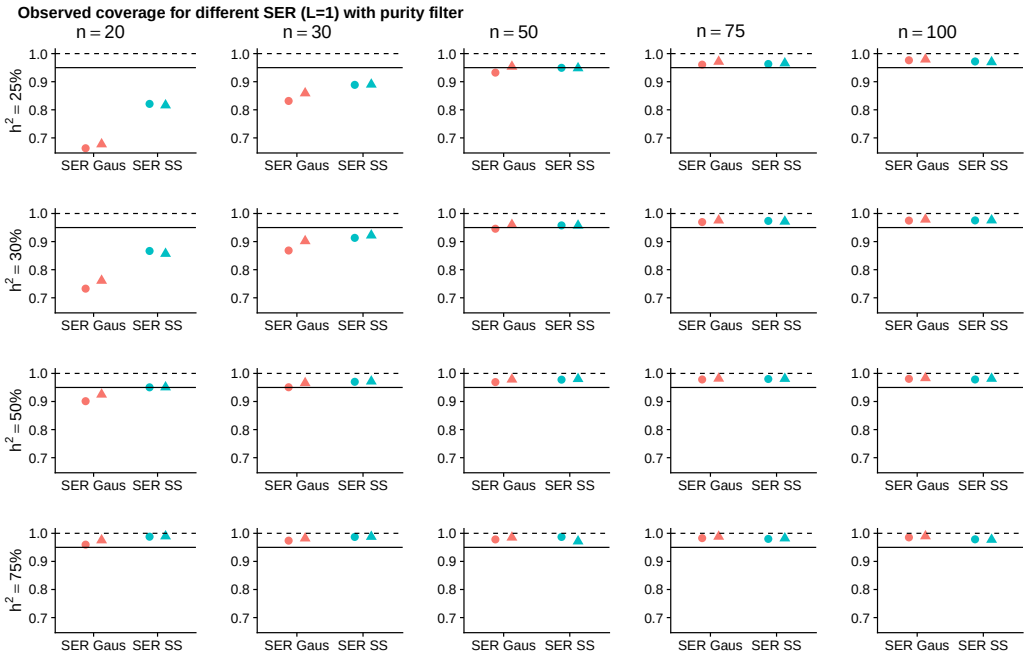

Fig. 7. Comparison of coverage between SuSiE using the default SER (red) and SuSiE using the SS-SER (blue) under a single causal SNP scenario with purity filtering. The solid line denotes the target 95% coverage level, while the dashed line indicates 100% coverage. Round markers represent simulations where SuSiE dynamically estimates the number of causal effects (with an upper bound of $L = 10$), whereas triangle markers correspond to simulations with the L parameter fixed to 1.

Methods

Simulations

In all simulations, we randomly sample a genomic region from the 1000 Genomes Phase 3 dataset [16] and select N unrelated individuals within this locus. We retain only biallelic SNPs with a minor allele frequency (MAF) $\geq$5% among the selected individuals.

Multiple causal SNPs simulations. For each simulation, we vary the number of individuals as $N = 20, 30, 50, 75$, and 100 and set the heritability, defined as the variance explained by the genotype, to $h^2 = 25\%, 30\%, 50\%$, and 75%.

For a given sample size N and heritability h^2, we randomly select N individuals as described above. We then randomly determine the number of causal SNPs, L, where L is drawn uniformly from $\{1, \ldots, 10\}$. The phenotype for individual i is generated as: $y_i = \mu_i + \epsilon_i$, where $\mu_i = \sum_{l=1}^{L} x_{il}$. Where x_{il} denotes the genotype at the lth causal SNP for individual i, and ϵ_i is a Gaussian noise term. To ensure that the in-sample variance explained by the genotype matches the specified h^2, we rescale the error term ϵ_i to satisfy the following identity:

$$h^2 = \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{3}$$

where $\bar{\mu}$ and $\bar{y}$ are the empirical means of μ and y, respectively. In these simulations, we set the upper bound for the number of effects to $L = 10$ when running SuSiE, regardless of the selected single-effect regression (SER) variant, and allow the method to estimate L.

Code availability. The scripts to replicate these simulations are available at `https://github.com/william-denault/susie_small_sample/`.

Single causal SNP simulations. Simulations for the single causal SNP scenario follow the same setup as in the multiple causal SNP scenario, except that we fix the number of causal SNPs at $L = 1$. For each combination of N and h^2, we evaluate the performance of SuSiE and its variants using the following configurations:

1. SuSiE with default SER, setting the upper bound on L to 1, with and without purity filtering.
2. SuSiE with default SER, setting the upper bound on L to 10, with and without purity filtering.
3. SuSiE with SS-SER, setting the upper bound on L to 1, with and without purity filtering.
4. SuSiE with SS-SER, setting the upper bound on L to 10, with and without purity filtering.

Mathematical expressions for the SS-SER model

Univariate regression model. For simplicity, we present all derivations assuming κ and λ are strictly positive, ensuring a well-defined prior. However, the quantities can be computed with $\kappa = 0, \lambda = 0$. Consider the following univariate regression model where σ^2 is treated as unknown:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}\beta + \epsilon \\ \epsilon &\sim N(0, \sigma^2) \\ \beta &\sim N(0, \sigma_0^2\sigma^2) \\ \sigma^2 &\sim \mathrm{IG}(\kappa/2, \lambda/2) \end{aligned} \tag{4}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are in $\in \mathbb{R}^n$, β is an unknown parameter, ϵ is Gaussian noise, and σ_0^2 is a scaling factor on the effect size. $\mathrm{IG}(\kappa, \lambda)$ denotes the inverse gamma distribution with shape κ and scale λ.

The posterior distribution of β conditional on the other inputs and σ^2 is

$$\beta \mid \boldsymbol{x}, \boldsymbol{y}, \sigma^2, \sigma_0^2 \sim N(\mu_1, \sigma^2 s_1^2), \tag{5}$$

where

$$\begin{aligned} s_1^2 &= \frac{r_0^2}{\boldsymbol{x}^T\boldsymbol{x}}, \\ \mu_1 &= r_0^2 \hat{b}, \\ r_0^2 &:= \frac{\sigma_0^2}{\sigma_0^2 + 1/(\boldsymbol{x}^T\boldsymbol{x})}. \end{aligned} \tag{6}$$

The posterior distribution of σ^2 is

$$\sigma^2 \mid \boldsymbol{x}, \boldsymbol{y}, \sigma_0^2 \sim \mathrm{IG}\left(\frac{\kappa_1}{2}, \frac{\lambda_1}{2}\right), \tag{7}$$

where

$$\begin{aligned} \kappa_1 &= \kappa + n, \\ \lambda_1 &= \lambda + \mathrm{RSS}(\sigma_0^2), \\ \mathrm{RSS}(\sigma_0^2) &= \boldsymbol{y}^T\boldsymbol{y}\left(1 - r_0^2\sigma_{xy}^2\right), \\ \sigma_{xy} &= \frac{\boldsymbol{x}^T\boldsymbol{y}}{\sqrt{(\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{y}^T\boldsymbol{y})}}. \end{aligned} \tag{8}$$

Thus, the posterior distribution of β is

$$\beta \mid \boldsymbol{x}, \boldsymbol{y}, \sigma_0^2 \sim t_{\kappa_1}\left(\mu_1, \frac{\lambda_1}{\kappa_1} s_1^2\right). \tag{9}$$

The Bayes factor (BF) for this model is given by

$$\mathrm{BF} = \sqrt{1 - r_0^2} \times \left(\frac{\lambda_1}{\lambda + \boldsymbol{y}^T\boldsymbol{y}}\right)^{-\kappa_1/2}. \tag{10}$$

Further details of these derivations are provided in the Supporting derivations.

Bayes factor and posterior distribution for SS-SER model. Under the SS-SER model (1–2), the posterior distribution of $\boldsymbol{\beta}_j$ is a mixture of a point mass at zero and a t-distribution:

$$\boldsymbol{\beta}_j | \boldsymbol{X}, \boldsymbol{y}, \sigma_0^2 = (1 - \pi_{1j})\delta_0 + \pi_{1j} t_{\kappa_{1j}}\left(\mu_{1j}, \frac{\lambda_{1j}}{\kappa_{1j}} s_{1j}^2\right) \tag{11}$$

where $s_{1j}^2, \mu_{1j}, \kappa_{1j}$, and λ_{1j} follow the same definitions as in the univariate case, substituting $\boldsymbol{x}$ with $\boldsymbol{X}_j$ in equation (4). Here, $y = \boldsymbol{X}_j\beta_j + \epsilon$. The mixture weights are given by:

$$\pi_{1j} = \frac{\pi_j \mathrm{BF}_j}{\sum_{j'=1}^{p} \pi_{j'} \mathrm{BF}_{j'}}, \tag{12}$$

with where BF_j is the Bayes factor for the univariate association between column j of $\boldsymbol{X}$ and $\boldsymbol{y}$ (see eq. 10).