

Realized Genome Sharing in Heritability Estimation Using Random Effects Models

Bowen Wang^{*,†} and Elizabeth Thompson^{*,1}

^{*}Department of Statistics, University of Washington, Seattle, Washington 98195-4322 and [†]Adobe Inc., San Jose, California 95110-2704

ORCID ID: 0000-0003-0198-7129 (E.T.)

ABSTRACT For heritability estimation using a two-component random effects model, we provided formulas for the limiting distribution of the maximum likelihood estimate. These formulas are applicable even when the wrong measure of kinship is used to capture additive genetic correlation. When the model is correctly specified, we showed that the asymptotic sampling variance of heritability estimate is determined by both the study design and the extent of variation in the kinship measure that constitutes the additive genetic correlation matrix. When the correlation matrix is mis-specified, the extent of asymptotic bias depends additionally on how the fitted correlation matrix differs from the truth. In particular, we showed in a simulation study that estimating heritability using a population-based design and the classic GRM as the fitted correlation matrix can potentially contribute to the “missing heritability” problem.

KEYWORDS

kinship
random effects
asymptotic bias
missing
heritability
model mis-specification

DNA inherited from the same ancestral copy by related individuals is said to be identical by descent (IBD). In the case of defined pedigrees, we measure IBD relative to pedigree founders. In quantitative genetics, correlation of trait values between relatives has often been modeled as a function of genome shared IBD. A useful parameter for pairwise IBD is the kinship coefficient, the probability that DNA randomly chosen from each individual at the same locus is IBD. Pedigree kinship (Ψ), the expectation of kinship coefficient over realizations of descent in the pedigree, is a deterministic function of the pedigree relationship. However, genome-wide realized kinship (Φ), the average of kinship coefficients over all loci in the genome varies widely around this expected value (Hill and Weir 2011). Advances in genetic marker data technology and statistical methodology have enabled us to estimate both local and genome-wide IBD sharing very accurately (e.g., Thompson 2013; Wang *et al.* 2017, and references therein). It is important to understand how differences in modeling genetic correlation between individuals affect outcome of statistical analyses.

We investigate the problem of heritability estimation using a random effects model of two components, where the trait values are sums of normally distributed additive genetic random effects and unique environmental random effects. The correlation structure of the additive genetic random effects is twice the matrix of some kinship measure. This model was used by Thompson and Shaw (1990) to show how eigen-transformation can significantly improve efficiency of EM algorithm to find the maximum likelihood estimates (MLEs) of the variance parameters. More recently, Visscher and Goddard (2015) used the same model to study asymptotic sampling variance (ASV) of heritability estimates. Raffa and Thompson (2016) considered hypothesis testing and construction of confidence intervals for heritability. These authors focused on the use of pedigree kinship (Ψ) under the assumption of correct model specification.

On the applied side, it has become popular to estimate heritability from population samples, where the genetic correlation structure takes the form of the classic Genomic Relationship Matrix (GRM). A population-based design avoids confounding due to shared environmental effects in close relatives. The classic GRM is constructed from identity by state (IBS) matching at genotyped markers, and is often referred to as the relatedness matrix (e.g., Yang *et al.* 2010). Wang *et al.* (2017) showed that the classic GRM (at the element level) is an unbiased (under mild assumptions), but not very accurate estimator of twice the genome-wide realized kinship, especially for remote relatives. It is of interest to know how choices of kinship measures (e.g., realized or pedigree) impact outcome of heritability estimation.

In this paper we study the asymptotic distributions of heritability estimates under the random effects model, when the genetic correlation

Copyright © 2019 Wang, Thompson

doi: <https://doi.org/10.1534/g3.119.0005>

Manuscript received January 4, 2019; accepted for publication March 15, 2019; published Early Online March 22, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

¹Corresponding author: Department of Statistics, Padelford Hall, University of Washington, Box 354322 Seattle, WA 98195-4322. E-mail: eathomp@uw.edu

structure is potentially mis-specified. We investigate how impact of mis-specifying genetic correlation structure on heritability estimation varies with respect to study design and the extent of variation in realized kinship. Finally, we discuss the use of estimated kinship to capture genetic correlation in a population-based design, and how it can contribute to the “missing heritability problem” (Maher 2008; Manolio *et al.* 2009).

METHODS

The polygenic model

In quantitative genetics, phenotype values are often modeled using a linear mixed model, where fixed effects may include age, sex, or principle components derived from the matrix of SNP genotypes to account for population sub-structure. In this paper we assume the fixed effects have been correctly adjusted for, so that

$$\mathbf{y} = \mathbf{g} + \mathbf{e}, \quad (1)$$

where \mathbf{y} are trait values after adjustment of fixed effects, $\mathbf{g} \sim \mathcal{N}(0, \sigma_g^2 \mathbf{G})$ are additive genetic effects and $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2 \mathbf{I})$ are residuals. σ_g^2 and σ_e^2 are unknown variance parameters, and \mathbf{G} is the appropriate genetic correlation matrix. Total phenotypic variance is $\sigma^2 = \sigma_g^2 + \sigma_e^2$. The goal is to estimate heritability, $h^2 = \sigma_g^2 / \sigma^2$. We can parametrize the trait distribution in terms of $\boldsymbol{\theta} = (h^2, \sigma^2)$ as

$$\mathbf{y} \sim \mathcal{N}(0, [h^2 \mathbf{G} + (1 - h^2) \mathbf{I}] \sigma^2). \quad (2)$$

We assume throughout the paper that the model described in (1) and (2) is true, and that h^2 and σ^2 are both positive. However, the fitted correlation matrix \mathbf{G}_f may differ from the true correlation matrix \mathbf{G}_t . To investigate the asymptotic distributions of the MLEs and the effect of pedigree structure on the MLEs, we assume there are m mutually independent pedigrees of the same structure with finite pedigree size n .

Correct model specification

When $\mathbf{G}_f = \mathbf{G}_t$, it follows from likelihood theory that the MLEs are consistent. Let the eigen-decomposition of \mathbf{G}_f be $\mathbf{T} \mathbf{D} \mathbf{T}^T$, where \mathbf{T} is the orthogonal matrix of eigenvectors and \mathbf{D} is the diagonal matrix of eigenvalues of \mathbf{G}_f . The transformed trait

$$\mathbf{y}^* = \mathbf{T}^T \mathbf{y} \sim \mathcal{N}(0, [h^2 \mathbf{D} + (1 - h^2) \mathbf{I}] \sigma^2). \quad (3)$$

The covariance matrix $\text{Var}(\mathbf{y}^*)$ is diagonal, so that the log-likelihood function without the constant term is

$$\ell(h^2, \sigma^2; \mathbf{y}^*) = -\frac{1}{2} \left[N \ln(\sigma^2) + \sum_{i=1}^N \ln(h^2 \lambda_i + 1 - h^2) + \frac{1}{\sigma^2} \sum_{i=1}^N \frac{y_i^{*2}}{h^2 \lambda_i + 1 - h^2} \right], \quad (4)$$

where $N = mn$ is the total sample size, λ_i is the i th eigenvalue of \mathbf{G}_f (and \mathbf{G}_t in the case). $\text{ASV}(\hat{h}^2)$ is given by, among others, Visscher and Goddard (2015) as

$$\text{ASV}(\hat{h}^2) = \frac{2}{\sum_{i=1}^N \frac{(\lambda_i - 1)^2}{(h^2 \lambda_i + 1 - h^2)^2} - \frac{1}{N} \left[\sum_{i=1}^N \frac{\lambda_i - 1}{h^2 \lambda_i + 1 - h^2} \right]^2}. \quad (5)$$

Model mis-specification

When $\mathbf{G}_f \neq \mathbf{G}_t$, let the matrix of differences be $\boldsymbol{\Delta} = \mathbf{G}_t - \mathbf{G}_f$. The same eigen-decomposition and transformation based on the fitted

correlation matrix \mathbf{G}_f leads to the same log-likelihood function (4), but the true distribution of the transformed trait is now

$$\mathbf{y}^* = \mathbf{T}^T \mathbf{y} \sim \mathcal{N}(0, [h_0^2 \mathbf{D} + (1 - h_0^2) \mathbf{I}] \sigma_0^2 + h_0^2 \sigma_0^2 \mathbf{T}^T \boldsymbol{\Delta} \mathbf{T}), \quad (6)$$

where $\boldsymbol{\theta}_0 = (h_0^2, \sigma_0^2)$ are the true parameter values.

Let $P(\boldsymbol{\theta}_0)$ denote the true trait distribution in (6). The model space for fitting with \mathbf{G}_f is denoted by $\mathcal{Q} = \{Q(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta = [0, 1] \times [0, \infty)\}$. If we let $m \rightarrow \infty$, the MLEs from fitting the wrong model, $\hat{\boldsymbol{\theta}} = (\hat{h}^2, \hat{\sigma}^2)$, will converge in probability to $\boldsymbol{\theta}_1 = (h_1^2, \sigma_1^2)$ that minimizes the Kullback-Leibler divergence (Kullback and Leibler 1951) between $P(\boldsymbol{\theta}_0)$ and $Q(\boldsymbol{\theta})$ over the parameter space Θ (e.g., Severini 2000). In our case

$$(h_1^2, \sigma_1^2) \approx \arg \min_{(h^2, \sigma^2)} \frac{1}{2} \ln(\sigma^2) + \frac{1}{2N} \sum_{i=1}^N \left[\ln(h^2 \lambda_i + 1 - h^2) + \frac{(h_0^2 \lambda_i + 1 - h_0^2) \sigma_0^2 + h_0^2 \sigma_0^2 \mathbf{t}_i^T \boldsymbol{\Delta} \mathbf{t}_i}{(h^2 \lambda_i + 1 - h^2) \sigma^2} \right], \quad (7)$$

where \mathbf{t}_i is the i th eigenvector of \mathbf{G}_f . It follows from likelihood theory that

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1) \rightarrow_d \mathcal{N}(0, \mathbf{J}^{-1} \mathbf{K} (\mathbf{J}^{-1})^T), \quad (8)$$

with

$$\mathbf{J} = \lim_{m \rightarrow \infty} \mathbb{E}_{P(\boldsymbol{\theta}_0)} \left[\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{y}^*) / N}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_1} \right], \quad (9)$$

$$\mathbf{K} = \lim_{m \rightarrow \infty} \mathbb{E}_{P(\boldsymbol{\theta}_0)} \left[\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}^*) / N}{\partial \boldsymbol{\theta}} \times \left(\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}^*) / N}{\partial \boldsymbol{\theta}} \right)^T \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_1} \right], \quad (10)$$

where $\ell(\boldsymbol{\theta}; \mathbf{y}^*)$ is the log-likelihood over all data. For instance, the first element of \mathbf{J} before taking limit can be shown to be

$$\mathbb{E}_{P(h_0^2, \sigma_0^2)} \left[\frac{\partial^2 \ell(h^2, \sigma^2; \mathbf{y}^*) / N}{\partial h^4} \Big|_{h_1^2, \sigma_1^2} \right] = \frac{1}{2N} \sum_{i=1}^N \left[\frac{(\lambda_i - 1)^2}{(h_1^2 \lambda_i + 1 - h_1^2)^2} - \frac{2(\lambda_i - 1)^2 [(h_0^2 \lambda_i + 1 - h_0^2) \sigma_0^2 + h_0^2 \sigma_0^2 \mathbf{t}_i^T \boldsymbol{\Delta} \mathbf{t}_i]}{\sigma_1^2 (h_1^2 \lambda_i + 1 - h_1^2)^3} \right]. \quad (11)$$

Given $h_0^2, \sigma_0^2, \mathbf{G}_t$ and \mathbf{G}_f for any finite $N = mn$, (7) can be solved numerically for (h_1^2, σ_1^2) . \mathbf{J} and \mathbf{K} in (9) and (10) can be computed subsequently to obtain the asymptotic variance covariance matrix of the MLEs.

For the special case where \mathbf{G}_f is twice the pedigree kinship matrix (2Ψ) and \mathbf{G}_t corresponds to realized kinship over the causal genome, elements of $\boldsymbol{\Delta}$ represent (twice the) deviations of realized kinship from its pedigree expectation. Such deviations have expectation 0 when samples are not ascertained by trait or by IBD sharing. Since the pedigree kinship matrix is fixed given the pedigree structure, we have $2\Psi_i = \sum_{j=1}^n \mathbf{t}_j \lambda_j \mathbf{t}_j^T$ for all i (pedigree index). It follows from (6) that $\boldsymbol{\Delta}$ enters (7), (9) and (10) only in the form of $\mathbf{t}_j^T (\sum \boldsymbol{\Delta}_i / m) \mathbf{t}_j$, which converges to 0 as $m \rightarrow \infty$. This implies fitting model (1) with $\mathbf{G}_f = 2\Psi$, regardless what is the true causal genome from which realized kinship is measured and captured in \mathbf{G}_t , produces consistent MLEs. In addition, the ASV of the MLEs will be the same as in the case $\mathbf{G}_f = \mathbf{G}_t = 2\Psi$, which can be easily computed using (5).

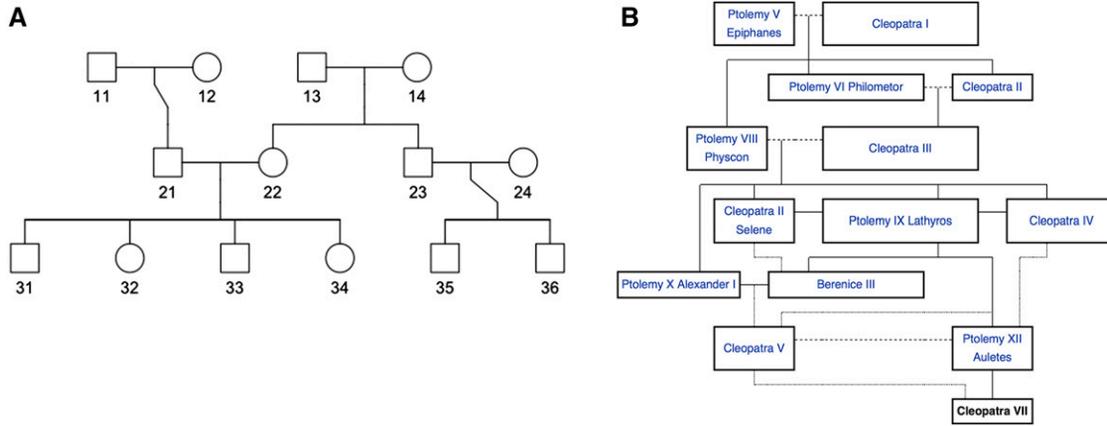


Figure 1 (A) Three-generation pedigree. (B) Cleopatra pedigree (Wikipedia 2018).

Method of analysis

We conduct two separate simulation studies for this paper. The first simulation study serves to verify results presented in the previous sections, and to investigate the impact of study design and the extent of variation in kinship measure on heritability estimation. We consider four different types of designs: sibpairs, sibships of 14, three-generation pedigree of 14 members (Figure 1A), and the Cleopatra pedigree of 14 members (Figure 1B). Three types of kinship measures are used for \mathbf{G}_t and \mathbf{G}_f : twice the pedigree kinship matrix (2Ψ), twice the genome-wide realized kinship matrix (2Φ), and twice the realized kinship matrix on chromosome 22 ($2\Phi^*$). These choices represent increasing amount of variation around the pedigree expectation.

We set total phenotypic variance $\sigma_0^2 = 1$ throughout, and use three heritability values, $h_0^2 \in \{0.2, 0.5, 0.8\}$. In total, there are 108 simulation scenarios ($4 \times$ designs, $9 \times (\mathbf{G}_t, \mathbf{G}_f)$ combinations, and $3 \times$ heritability values). The total sample size of the study is kept at $N = 1400$ (e.g., 700 sibpairs, or 100 sibships of 14). 500 simulation replicates are used to capture the empirical distributions of the MLEs in each of the 108 scenarios. Within each simulation replicate, we obtain \mathbf{G}_t and \mathbf{G}_f by simulating genome-wide descent in the pedigrees. Then we simulate trait data given \mathbf{G}_t and h_0^2 under model (1), and fit model (1) with \mathbf{G}_f to obtain the MLEs. Means and standard deviations of the empirical distribution of \hat{h}^2 are compared to those computed from (7) and (8) based on $(\mathbf{G}_t, \mathbf{G}_f)$.

The second simulation study investigates the potential impact of misspecifying the genetic correlation structure in a population-based design. For ease of simulation, we use multiple independent pedigrees of second or third cousinships of various sizes. This block-diagonal structure might approximate that of a population sample that included clusters of more closely related individuals. The total sample size is kept at 2000 across different simulation conditions, so that in one of the conditions, for example, the sample consists of 200 independent third cousinships of 10. We consider three different kinship measures, each computed within cousinship. These are twice the genome-wide realized kinship matrix (2Φ), the classic GRM ($2\Phi_c$), and the LD (linkage disequilibrium) weighted GRM ($2\Phi_w$) introduced in Wang *et al.* (2017). Both the classic GRM and the LD weighted GRM are computed from simulated marker data. The LD weighted GRM represents a more accurate estimator of twice the realized kinship matrix than the classic GRM. For the two GRMs, we look at the cases both with and without constraining the diagonal terms (twice the self-kinship) to 1. True parameter values are set to $h_0^2 = 0.5$ and $\sigma_0^2 = 1$.

Simulations were performed using the R package *rres* (Wang 2018). Genome-wide realized kinship between each pair of individuals was measured as the proportion of shared genome (in genetic distances) over the 22 autosomes. To simulate marker genotypes, real haplotypes from the 1000 Genomes Project Phase 3 data (The 1000 Genomes Project Consortium 2015) were assigned to founders and populated down the pedigree given the simulated joint IBD pattern. A total of 169,751 markers were selected from all 22 autosomes based on minor allele frequency, marker spacing, and availability of genetic position. Details of the marker data can be found in Wang *et al.* (2017). These SNP markers were used to obtain the classic GRM and the LD weighted GRM using the *rres* package, with allele frequencies assumed known.

Data availability

The 1000 Genomes Project Phase 3 data are available at <http://www.1000genomes.org/data>. Information on the set of 169,751 SNP markers used in the simulation studies can be found in the supplement of Wang *et al.* (2017).

RESULTS

In the first simulation study, there is generally very good agreement between the empirical and analytical results. Table 1 displays the results from the 9 combinations of $(\mathbf{G}_t, \mathbf{G}_f)$, for the sibship design and $h_0^2 = 0.5$. We see that \hat{h}^2 appears to be unbiased when $\mathbf{G}_f = \mathbf{G}_t$, and when $\mathbf{G}_f = 2\Psi \neq \mathbf{G}_t$. In addition, $\mathbf{G}_f = 2\Psi$ resulted in very similar $SE(\hat{h}^2)$ regardless of \mathbf{G}_t , as predicted in the previous section. When $\mathbf{G}_f \neq \mathbf{G}_t$ and \mathbf{G}_f uses a more variable kinship measure than \mathbf{G}_t (in this case, realized kinship on a shorter causal genome), both the point estimates and sampling errors tend to be smaller than expected under correct model specification.

When $\mathbf{G}_f = \mathbf{G}_t$, equation (5) suggests that pedigree structure affects $ASV(\hat{h}^2)$ only through eigenvalues of \mathbf{G}_t . For large N , the first term in the denominator of (5),

$$\sum_{i=1}^N \frac{(\lambda_i - 1)^2}{(h_0^2 \lambda_i + 1 - h_0^2)^2}, \quad (12)$$

dominates the second term. Since the summand in (12) is non-negative, eigenvalues that lead to bigger summand values have a higher tendency to reduce $ASV(\hat{h}^2)$. Figure 2A suggests that extreme eigenvalues have biggest impact in reducing sampling variance. Figure 2B shows that twice the realized kinship matrices on shorter genomic segments

■ **Table 1** Comparison of limits and asymptotic sampling standard errors of heritability estimates obtained by fitting model (1) and by using (7) and (8), with data on 100 independent sibships of 14. True parameters are $(h_0^2, \sigma_0^2) = (0.5, 1)$. Empirical distribution of \hat{h}^2 is constructed from 500 simulation replicates

G_t		2Ψ			2Φ			$2\Phi^*$		
G_f		2Ψ	2Φ	$2\Phi^*$	2Ψ	2Φ	$2\Phi^*$	2Ψ	2Φ	$2\Phi^*$
\hat{h}^2	emp.	0.494	0.451	0.278	0.497	0.498	0.281	0.492	0.486	0.499
	analy.	0.500	0.453	0.279	0.500	0.500	0.282	0.498	0.490	0.500
$SE(\hat{h}^2)$	emp.	0.063	0.052	0.032	0.065	0.060	0.032	0.064	0.059	0.032
	analy.	0.067	0.055	0.032	0.067	0.061	0.032	0.067	0.061	0.033

tend to produce more extreme eigenvalues. This provides intuition on why $SE(\hat{h}^2)$ is much smaller when $G_f = G_t = 2\Phi^*$, as compared to the other cases of correct model specification shown in Table 1.

Figure 3 shows simulation results from use of other designs and heritability values. We omit scenarios where $G_t = 2\Psi$ since it cannot be true biologically. Most of the observations from earlier discussion of the sibship design also hold true for other designs. The one clear exception is that when $G_t = 2\Phi$ and $G_f = 2\Psi$, the impact of this specific model mis-specification on $SE(\hat{h}^2)$ is very small when using other designs (compare blue circles to black bars in each scenario in Figure 3C). This is likely because the distribution of eigenvalues of the correlation matrix depends on the length of the causal genome as well as pedigree structure. For sibship design, a good proportion of eigenvalues associated with 2Φ are close to 0 (Figure 2B), which is not the case for other designs (results not shown). As a contrast, when $G_t = 2\Phi^*$ and $G_f = 2\Psi$, the impact of model mis-specification on $SE(\hat{h}^2)$ is clear for all designs (compare blue circles to black bars in each scenario in Figure 3D). This is because the greater variation in $2\Phi^*$ due to shorter length of causal genome leads to the possibility of more extreme eigenvalues for each study design.

Lastly, we note that for any given combination of (G_t, G_f) , $SE(\hat{h}^2)$ varies with h_0^2 quite differently across designs. For example, when $G_f = G_t = 2\Phi$ the sibship design has slightly smaller $SE(\hat{h}^2)$ than the three-generation design at $h_0^2 = 0.2$, but it is the opposite when $h_0^2 = 0.8$.

In the second simulation study, there is generally more bias in heritability estimates when the sample consists of third cousinships as opposed to second cousinships (Figure 4). This is in good agreement with the findings of Wang *et al.* (2017), where the authors showed it is relatively more difficult to estimate realized relatedness between more remote relatives. This observed association between level of bias and remoteness of relationship is further verified by additional simulations using full sibships, half sibships and first cousinships (results not shown).

Another observation from Figure 4 is that the decision to constrain the diagonal terms of $2\hat{\Phi}_c$ or $2\hat{\Phi}_w$ to 1 makes a big difference when pedigree sizes are small, but that difference erodes when pedigree sizes increase. A possible explanation is that apart from its effect on eigenvalues and eigenvectors in (7), $G_f \neq G_t$ induces bias in heritability estimates most directly from the matrix of difference, Δ . When we constrain the diagonal terms of $2\hat{\Phi}_c$ or $2\hat{\Phi}_w$ to 1, Δ will have 0's on the diagonal. This is expected to have a bigger impact when pedigree sizes are small and the diagonal terms make up a bigger proportion of non-zero terms of Δ .

Figure 4A and 4B shows that when $G_t = 2\Phi$, setting $G_f = 2\hat{\Phi}_w$ produces less bias than $G_f = 2\hat{\Phi}_c$ in all cases. This is expected since the LD weighted GRM is a more accurate estimator of twice the genome-wide realized kinship matrix. When constraining the diagonal terms of the estimated matrices to 1, bias increases with pedigree size for both second and third cousinships. Again, this is likely the result of the composition of non-zero terms in Δ : larger pedigree sizes imply higher proportion of non-zero terms. When not constraining the diagonal terms of the estimated matrices to 1, bias first decreases and then increases with pedigree size. A possible explanation is that in smaller cousinships, the diagonal terms of Δ take up a bigger proportion of all non-zero terms. While self-kinship can be estimated relatively accurately compare to realized kinship between remote cousins, the magnitude of deviations on the original scale, as captured in Δ , is greater for self-kinship (Wang *et al.* 2017). This means the diagonal terms of Δ could be more influential than off-diagonal terms on a per-entry basis, which lead to the bigger bias observed for smaller pedigrees.

Figure 4C and 4D show limits of heritability estimates when $G_t = 2\hat{\Phi}_c$, and $G_f = 2\Phi$ or $2\hat{\Phi}_w$. In both figures, the main observation is that the scale of bias is very small compare to that in Figure 4A and 4B. This matches our findings from the first simulation study: when G_t and G_f are associated with two kinship measures with zero expected differences (expectation over realization of descent on pedigree),

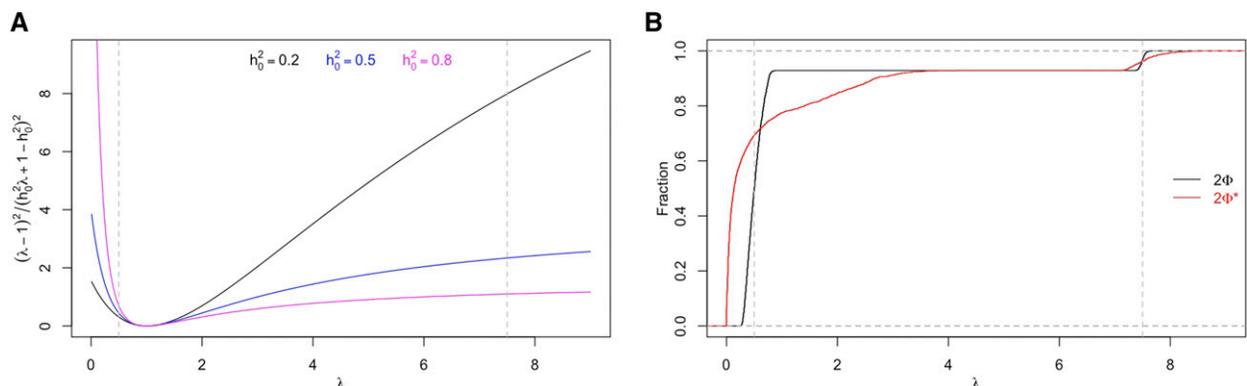


Figure 2 (A) $(\lambda - 1)^2 / (h_0^2 \lambda + 1 - h_0^2)^2$ evaluated at a range of λ values and $h_0 \in \{0.2, 0.5, 0.8\}$. (B) Cumulative distributions of eigenvalues of 2Φ and $2\Phi^*$ from 100 simulated sibships of 14. Vertical dashed lines in both plots represent the two distinct eigenvalues of 2Ψ : 0.5 and 7.5.

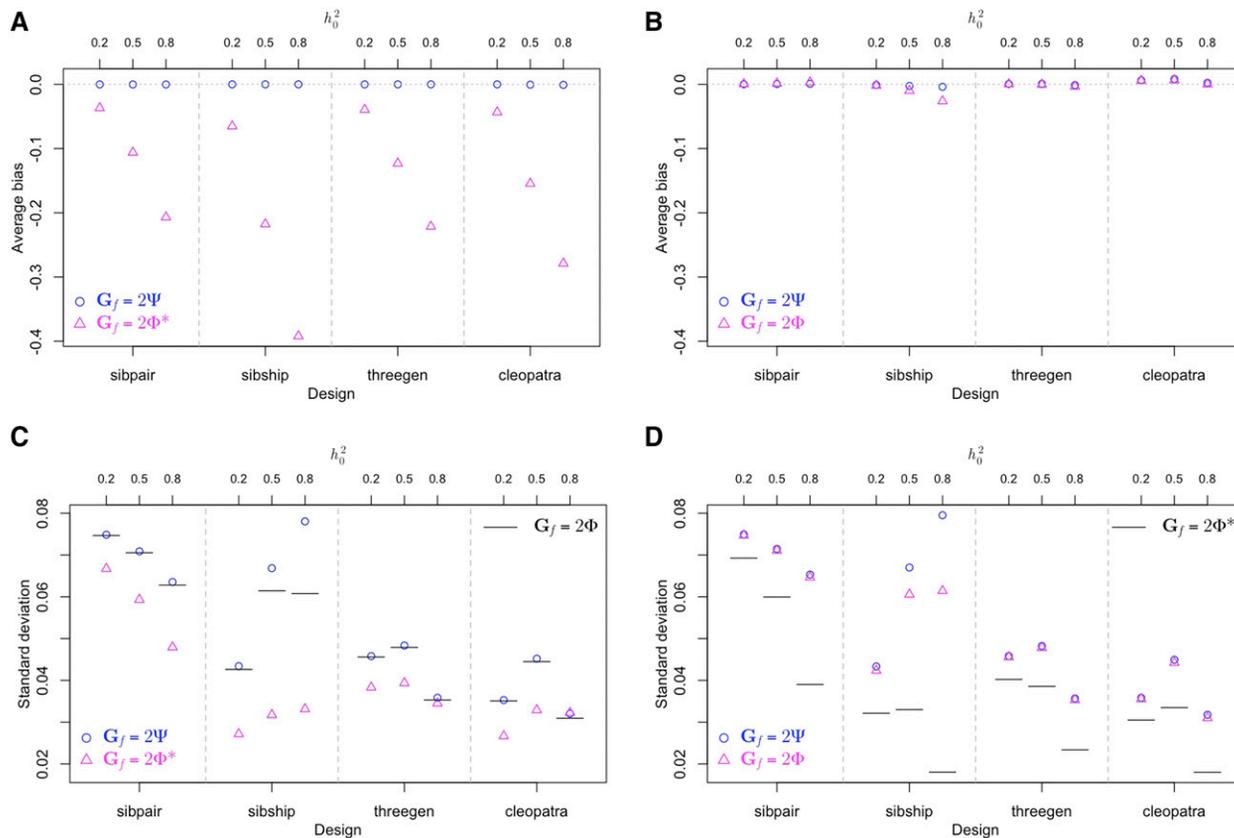


Figure 3 Simulation results for all four designs, three values of h_0^2 and various combinations of (G_t, G_f) . (A) and (B) show the average bias of heritability estimates from 500 simulation replicates, when $G_t = 2\Phi$ or $2\Phi^*$ respectively. (C) and (D) show the standard deviation of heritability estimates from 500 simulation replicates, when $G_t = 2\Phi$ or $2\Phi^*$ respectively. In both (C) and (D), results obtained under correct model specification are shown in black bars as references.

we expect downward bias in heritability estimate if G_f uses a more variable kinship measure than G_t , but not much bias the other way round (e.g., Figure 3A and 3B). Wang *et al.* (2017) showed that both the classic GRM and the LD weighted GRM are unbiased estimators of twice the genome-wide realized kinship under certain assumptions, and the unbiasedness property held very well in simulation studies. Since the classic GRM is a less efficient estimator, it is a more variable measure of kinship compare to LD weighted GRM.

DISCUSSION

For heritability estimation using a two-component random effects model, we have provided formulas for the limits and the asymptotic sampling variances of the MLEs. These formulas are applicable even when the wrong kinship measure is used to capture additive genetic correlation. To study the impact of pedigree structure on asymptotic distribution of the MLEs, we have assumed that the sample consists of multiple independent pedigrees of the same structure. The assumption of a fixed pedigree structure is not required to derive equations (4) through (11) in **METHODS**, since log-likelihood sums over independent pedigrees of any structure. When different pedigree structures are present in the study sample, one only needs to replace the fixed pedigree size n with n_i for pedigree i in those formulas and the asymptotic results are still valid.

Under correct specification of model (1), ASV of heritability estimates is a function of eigenvalues of the appropriate genetic correlation matrix, which in turn can depend on pedigree structure and variation in the corresponding kinship measure. For any pedigree structure and

choice of kinship measure, the resulting kinship matrix can be easily obtained by simulation and subsequently used in the formulas we presented to assess effectiveness of the pedigree in heritability estimation under the assumed model. This is beneficial to pedigree selection in study design.

There can be bias in heritability estimate if the genetic correlation matrix has been mis-specified. The extent of the bias depends on both the true and the fitted correlation matrices, as well as the study design. When the true genetic correlation between individuals is captured by genome-wide realized kinship, one should use an accurate estimator of realized kinship to compute the fitted correlation matrix in order to reduce downward bias in heritability estimate. This is especially important when the study sample contains many remotely related individuals.

The popular classic GRM is not an accurate estimator of twice the realized kinship. When used in a population-based design, it can lead to substantial downward bias in heritability estimate if the truth (G_t) is that of twice the genome-wide realized kinship. The downward bias is expected to persist even if denser SNP panels are used to estimate realized kinship, as additional markers do not provide information without limit. In a follow up study of Wang *et al.* (2017), we found that increasing marker density by 4 times made little improvement in accuracy of the classic GRM estimator (results not shown). This choice of kinship estimator and study design can contribute to the “missing heritability” problem, among many other factors. On the other hand, using a more accurate kinship estimator to compute the fitted correlation matrix (G_f) is more robust to mis-specification of kinship measure.

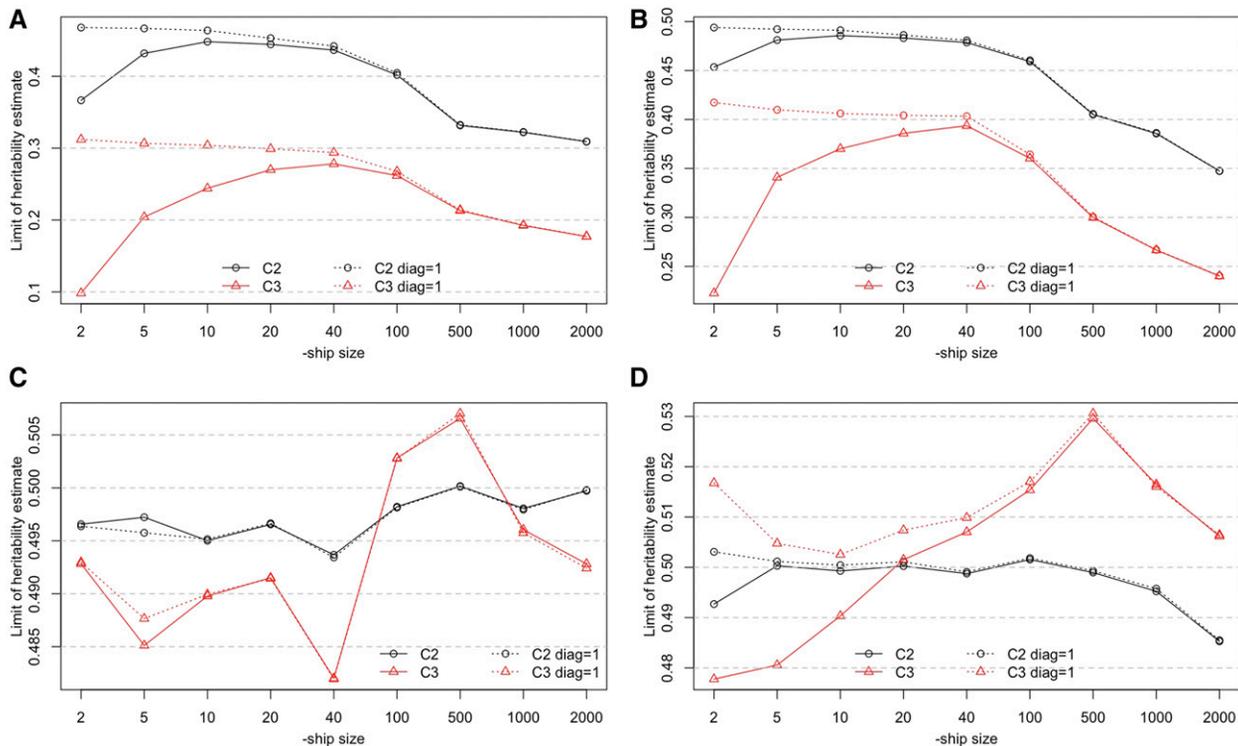


Figure 4 Limits of \hat{h}^2 when $h_0^2 = 0.5$ and $G_f \neq G_t$. Each sample contains multiple independent second (C2) or third (C3) cousinships of specific sizes. Three kinship measures considered are: twice the genome-wide realized kinship matrix (2Φ), the classic GRM ($2\Phi_c$) and the LD weighted GRM ($2\Phi_w$). In (A) and (B), $G_t = 2\Phi$, $G_f = 2\Phi_c$ and $2\Phi_w$ respectively. In (C) and (D), $G_t = 2\Phi_c$, $G_f = 2\Phi$ and $2\Phi_w$ respectively. The estimated matrices ($2\Phi_c$ and $2\Phi_w$) are used with and without constraining the diagonal terms to 1.

When using a population-based design, it is common to remove an individual in each pair that has a kinship estimate exceeding certain threshold (e.g., Yang *et al.* 2010). A threshold of 0.025 roughly corresponds to excluding closer relatives than second cousins (with pedigree kinship 0.0156). While more remote relatives often exist in a population sample, we only presented results on second and third cousinships in the second simulation study because they demonstrate all the important points from this simulation setup. The main reason for using a population-based design is not to eliminate IBD, but to avoid confounding due to shared environmental effects of closer relatives. Since our simulation models do not include shared environmental effects, our findings should generalize to studies that involve more remote relatives than third cousins. Additional results on full sibships, half sibships and first cousinships (not shown) have all confirmed the findings we have discussed. In particular, asymptotic bias in heritability increases with remoteness of relationship when the additive genetic correlation matrix is mis-specified.

Formulas in **METHODS** do not associate G_t or G_f with specific kinship measures. They are applicable to all choices of genetic correlation matrices. The decision to assume realized *vs.* estimated kinship as the appropriate kinship measure for G_t depends on one's belief on whether genetic correlation is best captured by IBD sharing or by IBS matching at genotyped markers. Taking the IBS perspective, Jiang *et al.* (2016) provided a theoretical justification for consistency of heritability estimates in genome-wide association studies (GWAS) under mis-specified linear mixed model. The authors assumed that G_t takes the form of the classic GRM, but the fitted model is mis-specified in the sense that only a subset of the GWAS SNPs are actually causal. This setup falls within the more

general discussions in this paper. We have shown that fitting a covariance matrix that is broader than truth causes little asymptotic bias (e.g., Figure 3B).

A model with only two random effects is attractive for ease of computation and interpretability. In particular, an eigen-transformation is possible so that variance covariance matrix of the transformed trait is diagonal, and it depends on the appropriate genetic correlation matrix only through the eigenvalues. This simplification is not possible with more than two random effects. The asymptotic properties of heritability estimates under more complicated random effects models will be a topic of future investigation.

ACKNOWLEDGMENTS

We thank the two reviewers for their constructive comments. This research was supported in part by National Institutes of Health grants R37 GM-046255.

LITERATURE CITED

- Hill, W. G., and B. S. Weir, 2011 Variation in actual relationship as a consequence of mendelian sampling and linkage. *Genet. Res.* 93: 47–64. <https://doi.org/10.1017/S0016672310000480>
- Jiang, J., C. Li, D. Paul, C. Yang, and H. Zhao, 2016 On high-dimensional misspecified mixed model analysis in genome-wide association study. *Ann. Stat.* 44: 2127–2160. <https://doi.org/10.1214/15-AOS1421>
- Kullback, S., and R. A. Leibler, 1951 On information and sufficiency. *Ann. Math. Stat.* 22: 79–86. <https://doi.org/10.1214/aoms/1177729694>
- Maher, B., 2008 Personal genomes: The case of the missing heritability. *Nature* 456: 18–21. <https://doi.org/10.1038/456018a>
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* 461: 747–753. <https://doi.org/10.1038/nature08494>

- Raffa, J. D., and E. A. Thompson, 2016 Power and effective study size in heritability studies. *Stat. Biosci.* 8: 264–283. <https://doi.org/10.1007/s12561-016-9143-2>
- Severini, T. A., 2000 *Likelihood Methods in Statistics*, Oxford University Press, Oxford, UK.
- The 1000 Genomes Project Consortium, 2015 A global reference for human genetic variation. *Nature* 526: 68–74. <https://doi.org/10.1038/nature15393>
- Thompson, E. A., 2013 Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* 194: 301–326. <https://doi.org/10.1534/genetics.112.148825>
- Thompson, E. A., and R. G. Shaw, 1990 Pedigree analysis for quantitative traits: variance components without matrix inversion. *Biometrics* 46: 399–413. <https://doi.org/10.2307/2531445>
- Visscher, P. M., and M. E. Goddard, 2015 A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* 199: 223–232. <https://doi.org/10.1534/genetics.114.171017>
- Wang, B., 2018 *rres: Realized Relatedness Estimation and Simulation*. R package version 1.1. <https://cran.r-project.org/web/packages/rres>
- Wang, B., S. Sverdlov, and E. A. Thompson, 2017 Efficient estimation of realized kinship from single nucleotide polymorphism genotypes. *Genetics* 205: 1063–1078. <https://doi.org/10.1534/genetics.116.197004>
- Wikipedia, 2018 Cleopatra pedigree. <http://en.wikipedia.org/wiki/Cleopatra#Ancestry>.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common snps explain a large proportion of the heritability for human height. *Nat. Genet.* 42: 565–569. <https://doi.org/10.1038/ng.608>

Communicating editor: R. W. Doerge