

A General Unified Framework to Assess the Sampling Variance of Heritability Estimates Using Pedigree or Marker-Based Relationships

Peter M. Visscher^{*,†,1} and Michael E. Goddard^{*,§}

^{*}Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia [†]The University of Queensland Diamantina Institute, The Translational Research Institute, Brisbane, Queensland 4102, Australia, [‡]University of Melbourne, Department of Food and Agricultural Systems, Parkville, Victoria 3010, Australia, and [§]Biosciences Research Division, Department of Environment and Primary Industries, Bundoora, Victoria 3001, Australia

ABSTRACT Heritability is a population parameter of importance in evolution, plant and animal breeding, and human medical genetics. It can be estimated using pedigree designs and, more recently, using relationships estimated from markers. We derive the sampling variance of the estimate of heritability for a wide range of experimental designs, assuming that estimation is by maximum likelihood and that the resemblance between relatives is solely due to additive genetic variation. We show that well-known results for balanced designs are special cases of a more general unified framework. For pedigree designs, the sampling variance is inversely proportional to the variance of relationship in the pedigree and it is proportional to $1/N$, whereas for population samples it is approximately proportional to $1/N^2$, where N is the sample size. Variation in relatedness is a key parameter in the quantification of the sampling variance of heritability. Consequently, the sampling variance is high for populations with large recent effective population size (e.g., humans) because this causes low variation in relationship. However, even using human population samples, low sampling variance is possible with high N .

HERITABILITY (h^2), the proportion of phenotypic variation that is explained by additive genetic variation, is an important parameter in plant and animal breeding, evolutionary genetics, and human and medical genetics. It is central in quantifying the role of genetics in complex traits, predicting response to selection in natural and artificial breeding programs, and determining the limits of trait or disease prediction using information from relatives or DNA markers. Traditionally, the estimation of heritability is from pedigree data, by modeling the observed resemblance between relatives (Falconer and Mackay 1996; Lynch and Walsh 1998). More recently, genetic variation has been estimated using genetic marker information (Ritland 2000; Thomas 2005; Visscher *et al.* 2006; Yang *et al.* 2010; Robinson *et al.* 2013; Berenos *et al.* 2014). These designs estimate the genetic variance explained by the markers, which may be less than the additive genetic variance (Yang *et al.* 2010), but in this article

we refer to the parameter estimated as the heritability regardless of whether it is estimated from relationships defined by pedigree or by markers. In general, designs to estimate heritability can be grouped by their use of (i) the expected identity-by-descent (IBD) sharing between relatives, *i.e.*, using pedigree relationships, (ii) marker-based estimated IBD relationships between relatives for known pedigree relationships, and (iii) marker-based estimated genomic relationship matrices for unknown pedigree relationships. For a review of these designs with a particular focus on human populations, see Vinkhuyzen *et al.* (2013).

Even with large sample sizes, the standard error of heritability estimates is often disappointingly large and it varies greatly between experimental designs. Therefore it is important to calculate the expected standard error before committing resources to collecting the data. Given a particular experimental design and the population value of h^2 , its sampling variance can be determined using a number of methods. After the data have been collected, the (asymptotic) sampling variance of the estimate can be derived from the analysis, for example, from mean squares in balanced designs, from the information matrix when using maximum likelihood or from the posterior density in Bayesian analysis.

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.114.171017

Manuscript received September 17, 2014; accepted for publication October 27, 2014;
published Early Online October 31, 2014.

Available freely online through the author-supported open access option.

¹Corresponding author: Queensland Brain Institute, University of Queensland, St. Lucia, Queensland 4072, Australia. E-mail: peter.visscher@uq.edu.au

Prior to collecting data on phenotypes, the sampling variance can be predicted using statistical theory, typically for balanced designs, or obtained from computer simulation for more complex pedigree structures. In this study, we provide a single framework for calculating the asymptotic sampling variance of the heritability across a wide range of designs, for a class of models with two random variables and when analysis is by maximum likelihood (ML). We derive the sampling variance using the expected value of the information matrix. We show that previous results are special cases of the general framework and that the variance in relationships in the sample is a key parameter in all experimental designs.

Model and Assumptions

We assume a linear model with no fixed effects (or fixed effects that have been adjusted for without error) and two random components, a genetic effect (\mathbf{g}), and a residual effect (\mathbf{e}). There are N individuals, each with a single observation, y ,

$$\mathbf{y} = \mathbf{g} + \mathbf{e}, \text{ with } \text{var}(\mathbf{g}) = \mathbf{G}\sigma_g^2 \text{ and } \text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2,$$

where \mathbf{y} , \mathbf{g} , and \mathbf{e} are vectors of length N of the phenotypic observations, genetic value, and residuals, respectively. \mathbf{G} is the genetic relationship matrix (GRM), either from pedigree relationships, in which case it is the usual numerator relationship matrix (twice the kinship matrix), or derived from SNP similarity (Vanraden 2008; Strandén and Garrick 2009; Yang *et al.* 2010). The genetic, residual, and total variances are σ_g^2 , σ_e^2 , and σ^2 , respectively. The $N \times N$ covariance matrix of all observations (\mathbf{V}) is

$$\text{var}(\mathbf{y}) = \mathbf{V} = [\mathbf{G}h^2 + \mathbf{I}(1 - h^2)]\sigma^2,$$

where $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2) = \sigma_g^2/\sigma^2$, the heritability.

General Formula for Sampling Variance

We can decompose the symmetric GRM as

$$\mathbf{G} = \mathbf{T}\mathbf{D}\mathbf{T}',$$

with $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ and $\mathbf{T}^{-1} = \mathbf{T}'$ because \mathbf{T} is orthogonal and \mathbf{D} a diagonal matrix containing eigenvalues (λ_i) of \mathbf{G} . Inference on h^2 from data \mathbf{y} does not change upon a linear transformation of \mathbf{y} . We can therefore transform \mathbf{y} by using the eigenvectors of \mathbf{G} , which for the simple model used here are also eigenvectors of \mathbf{V} (Thompson and Shaw 1990, 1992; Lippert *et al.* 2011; Blangero *et al.* 2013; Raffa and Thompson 2014).

Define $\mathbf{y}^* = \mathbf{T}^{-1}\mathbf{y} = \mathbf{T}'\mathbf{y}$. Then

$$\mathbf{y}^* = \mathbf{T}'\mathbf{g} + \mathbf{T}'\mathbf{e} = \mathbf{g}^* + \mathbf{e}^*,$$

with

$$\begin{aligned} \text{var}(\mathbf{y}^*) &= \mathbf{P} = \left[\mathbf{T}'\mathbf{G}\mathbf{T}h^2 + \mathbf{T}'\mathbf{T}(1 - h^2) \right] \sigma^2 \\ &= [\mathbf{D}h^2 + \mathbf{I}(1 - h^2)] \sigma^2. \end{aligned}$$

The log likelihood with respect to h^2 and σ^2 is

$$\begin{aligned} \log L &= -\frac{1}{2} \left[\log |P| + \mathbf{y}^{*'} P^{-1} \mathbf{y}^* \right] \\ &= -\frac{1}{2} \left[N \times \log(\sigma^2) + \sum \log(\lambda_i h^2 + 1 - h^2) \right. \\ &\quad \left. + (1/\sigma^2) \sum (y_i^{*2} / (\lambda_i h^2 + 1 - h^2)) \right], \end{aligned} \quad (1)$$

as shown previously (Thompson and Shaw 1990; Raffa and Thompson 2014). Equation 1 is very similar to that in Blangero *et al.* (2013), but with added parameter σ^2 . Elements of the (Fisher) information matrix (\mathbf{F}) are obtained by taking the second derivative of (1) taken at the maximum with respect to h^2 and σ^2 , and then the negative value of its expectation over \mathbf{y}^* , using

$$\begin{aligned} E(y_i^{*2}) &= \text{var}(y_i^*) = (\lambda_i h^2 + 1 - h^2) \sigma^2 \\ &= (1 + h^2(\lambda_i - 1)) \sigma^2. \end{aligned}$$

The derivation of the first element of \mathbf{F} (F_{11}) is given here. The other two elements are derived analogously,

$$\begin{aligned} \delta \log L / \delta h^2 &= -\frac{1}{2} \left[\sum (\lambda_i - 1) / (\lambda_i h^2 + 1 - h^2) \right. \\ &\quad \left. - (1/\sigma^2) \sum (y_i^{*2} (\lambda_i - 1) / (\lambda_i h^2 + 1 - h^2)^2) \right] \\ \delta \log L^2 / \delta h^4 &= -\frac{1}{2} \left[-\sum (\lambda_i - 1)^2 / (\lambda_i h^2 + 1 - h^2)^2 \right. \\ &\quad \left. + 2(1/\sigma^2) \sum (y_i^{*2} (\lambda_i - 1)^2 / (\lambda_i h^2 + 1 - h^2)^3) \right] \end{aligned}$$

and so

$$\begin{aligned} F_{11} &= -E(\delta \log L^2 / \delta h^4) \\ &= \frac{1}{2} \sum \left[-(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2 \right] \\ &\quad + 2(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2 \\ &= \frac{1}{2} \sum \left[(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2 \right]. \end{aligned}$$

The resulting elements of the 2×2 matrix \mathbf{F} are $F_{11} = \frac{1}{2}a$, $F_{12} = F_{21} = \frac{1}{2}b/\sigma^2$, $F_{22} = \frac{1}{2}N/\sigma^4$, with constants a and b ,

$$a = \sum \left[(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2 \right],$$

and

$$b = \sum \left[(\lambda_i - 1) / (1 + h^2(\lambda_i - 1)) \right].$$

These elements are similar to those presented in Thompson and Atkins (1994), who parameterized the likelihood in a genetic and residual variance component, whereas we have parameterized in heritability and phenotypic variance. Thompson and Atkins do not have the factor $\frac{1}{2}$ and have λ_i^2 and λ_i in the equations above where we have $(\lambda_i - 1)^2$ and $(\lambda_i - 1)$, respectively, the difference due to the choice of parameters in the model. In the article that developed the method of estimation of variance component in linear mixed models using restricted maximum likelihood (Patterson and Thompson 1971), the authors presented both the log likelihood and the information matrix in terms of eigenvalues of the covariance matrix.

The asymptotic sampling (co)variance for the estimates of heritability and phenotypic variance are from \mathbf{F}^{-1} . Therefore, the asymptotic sampling variance of the estimate of the heritability is

$$\text{var}(\hat{h}^2) \approx 2/(a - b^2/N). \quad (2)$$

Hence, under the assumptions given, this is a completely general expression for the asymptotic sampling variance of an estimate of heritability and depends only on the eigenvalues of the GRM, the population value of heritability, and the experimental sample size.

Special Cases

With additional assumptions or for balanced designs, terms for a and b simplify and simple solutions for the sampling variance of \hat{h}^2 can be derived. We go through a number of these special cases in this section that encompass pedigree and marker-based GRM.

Phenotypic variance (σ^2) known

In many applications, the sampling variance of the total phenotypic variance is small or known before the experiment is conducted, and therefore it is useful to consider the sampling variance of heritability under the assumption that the phenotypic variance is known without error. For example, Blangero *et al.* (2013) assume that σ^2 is known in their derivations of the expected likelihood-ratio-test statistic (ELRT). If we assume here that the phenotypic variance is known without error then the resulting sampling variance of the estimate of heritability is

$$\text{var}(\hat{h}^2 | \sigma^2 \text{ known}) = 2/a. \quad (3)$$

This expression is smaller than that in (2); hence assuming that phenotypic variance is known when it is not will lead to an underestimate of the sampling variance of heritability. This underestimate will be small when b^2/N is small relative to the term a .

$h^2 \rightarrow 0$

For a small heritability, $a \rightarrow \sum (\lambda_i - 1)^2$, $b \rightarrow \sum (\lambda_i - N)$, and

$$\text{var}(\hat{h}^2 | h^2 \rightarrow 0) \approx 2/[N \times \text{var}(\lambda_i)]. \quad (4)$$

Assuming that the phenotypic variance is known and h^2 is small gives

$$\text{var}(\hat{h}^2 | \sigma^2 \text{ known}, h^2 \rightarrow 0) \approx 2/[N \times \text{var}(\lambda_i) + N(E(\lambda_i) - 1)^2],$$

which is close to (4) because the mean eigenvalue will be 1 in the absence of inbreeding when the GRM is from pedigree identity-by-descent and very close to 1 when the GRM is estimated from SNP data (Janss *et al.* 2012). Hence, when the population value of heritability is small, its sampling variance is only a function of the variation in relatedness and sample size.

All $\lambda_i \rightarrow 1$

Equation 4 is also the result for when all λ_i are close to 1, such that their variance approaches zero. This situation can occur when the GRM is created from population SNP data on unrelated individuals in a population with a large effective population size. However, as we derive below, the variance of eigenvalues depends both on experimental sample size and effective population size, and so these parameters affect the sampling variance of heritability. In particular, the variance in eigenvalues is proportional to experimental sample size, so the larger the sample size the wider the spread around a mean value of 1.

Pairs of relatives with relationship r

If there are m pairs of relatives of the same degree r , then $2m = N$ and there are m eigenvalues λ_1 with value $1 + r$ and m eigenvalue λ_2 with value $1 - r$ (Searle 1982; Blangero *et al.* 2013). Let $\rho = rh^2$. Then

$$\begin{aligned} a &= 2mr^2(1 + r^2h^4) / (1 - r^2h^4)^2 \\ &= 2mr^2(1 + \rho^2) / (1 - \rho^2)^2, \end{aligned}$$

$$b = -2mr^2h^2 / (1 - r^2h^4) = -2mr\rho / (1 - \rho^2),$$

and

$$\text{var}(\hat{h}^2) = (1 - \rho^2)^2 / (mr^2). \quad (5)$$

For pairs of monozygotic (MZ) twins ($r = 1$), Equation 5 becomes $\text{var}(\hat{h}^2) = (1 - \rho^2)^2 / m$. For pairs of full-sibs ($r = \frac{1}{2}$), the sampling variance is $4(1 - \rho^2)^2 / m$. For bivariate normality, the sampling variance of a correlation coefficient between two variates with population value ρ is $\sim (1 - \rho^2)^2 / N$ (e.g., Lynch and Walsh 1998, p. 819), so consistent with Equation 5.

Balanced design of multiple families

For m families with n individuals of relationship r , there are $(n - 1)$ eigenvalues of $(1 - r)$ and 1 eigenvalue of $(1 + r(n - 1))$

per family. This follows from known results on eigenvalues for symmetrical matrices that can be written as $c\mathbf{I} + d\mathbf{J}$, with c and d constants (Searle 1982). Substituting these eigenvalues into the equation for parameters a and b gives

$$a = [m r^2 n(n-1)(1+\rho^2(n-1))] / [(1-\rho)^2(1+(n-1)\rho)]^2$$

$$b = -mn(n-1)r\rho / [(1-\rho)(1+\rho(n-1))]$$

and

$$\text{var}(\hat{h}^2) \approx 2(1-\rho)^2[1+(n-1)\rho]^2 / [mn(n-1)r^2]. \quad (6)$$

This is consistent with the intraclass correlation sampling variance (e.g., Falconer and Mackay 1996, p. 180), apart from having m in the denominator [the least-squares derivation has $(m-1)$ instead]. Although we have assumed no fixed effects, in practice at least a mean would be included in the model and this absorbs one degree of freedom from the comparison of families. The least-squares formula takes account of this but ML estimation ignores it. Assuming that the phenotypic variance is known gives

$$\text{var}(\hat{h}^2) \approx 2(1-\rho)^2[1+(n-1)\rho]^2 / [mn(n-1)r^2 \times (1+\rho^2(n-1))],$$

smaller than (5) by a factor of $1/(1+\rho^2(n-1))$. For large half-sib families, this term can be substantial.

Twin design

In human populations, the classical twin design is common for estimating genetic and nongenetic variance components. Let $N = 2m_M + 2m_D$, with m_M and m_D the number of MZ and dizygotic (DZ) pairs, respectively. In total, there are four different eigenvalues: 2, 0, 3/2, and 1/2 (Blangero *et al.* 2013), with multiplicity m_M , m_M , m_D , and m_D . Let $c = m_M/(m_M + m_D)$, the proportion of all twin pairs that are MZ pairs. Using Equation 5, $a - b^2/N = NT$, with

$$\begin{aligned} T &= c(1+h^4(1-c)) / (1-h^4)^2 \\ &+ \frac{1}{4}(1-c) \left(1 + \frac{1}{4}h^4c\right) / \left(1 - \frac{1}{4}h^4\right)^2 \\ &- \frac{1}{4}c(1-c)h^4 / \left[(1-h^4) \left(1 - \frac{1}{4}h^4\right)\right] \end{aligned}$$

and

$$\text{var}(\hat{h}^2) = (2/N)T^{-1}.$$

This analysis assumes that there are no common environmental effects so the sampling variance is not appropriate for the usual practice of estimation of heritability using maximum likelihood fitting both an additive genetic and common environmental component (Neale and Cardon 1992).

Within-family estimation using realized relationships estimates from markers

Full-sibs have an expected pedigree relationship of 0.5 but the actual amount of the genome shared varies around 0.5 and this realized relationship can be estimated using genetic markers and used to estimate heritability (Visscher *et al.* 2006, 2007; Hemani *et al.* 2013). These relationships can be estimated using identity-by-descent calculations conditional on observed marker genotypes. For full-sibs and half-sibs in human populations, the standard deviation of realized relationships is ~ 0.04 and 0.03 , around the expected value of $\frac{1}{2}$ and $\frac{1}{4}$, respectively. For a comprehensive theory on the variance of realized relationships, see Hill and Weir (2011). A feature of this design is that common environmental factors that vary between families do not bias the heritability estimate. Visscher *et al.* (2006) derived an approximate sampling variance of the estimate of heritability from multiple families with two full-sibs each. Hill (2013) derived the sampling variance of the estimate of genetic variation using REML for the general case of f families each of size n and expected relationship θ (twice the kinship coefficient). We can use the same general framework as developed here to approximate the sampling variance from within-family estimation. The difference between this design and those previously discussed is that the GRM is not fixed. That is, the eigenvalues of the GRM are themselves random variables and to derive the sampling variance of the estimate of heritability we need to first derive the expected value of the elements of the Information matrix over repeated samples. We provide details of an approximation in Appendix A. It results in

$$\begin{aligned} \text{var}(\hat{h}^2) &\approx \left[2(1-t)^2 / (f \times n^{2*} \text{var}(r_{ij})) \right] \\ &\times \left[(1-t)^2 - nh^4 \text{var}(r_{ij}) \right] \\ &= \left[2(1-t)^2 / (Nn \times \text{var}(r_{ij})) \right] \\ &\times \left[(1-t)^2 - nh^4 \text{var}(r_{ij}) \right]. \end{aligned} \quad (7)$$

This equation shows that the sampling variance reduces by the square of the sample size per family (n), essentially because every individual adds a contrast with all other family members in the sample. As detailed in Appendix A, this approximation breaks down when h^2 and n are large.

Random sampling from the population

One design to estimate the amount of additive genetic variation captured by SNPs is to take a random sample of individuals from the population, derive a GRM from SNP similarity, and estimate variance components from (residual) maximum likelihood (Yang *et al.* 2010). In this sampling scheme, individuals are not sampled or ascertained based upon particular pedigree relationships, and any pedigree relationship, if known, is not taken into account in the analysis. The sampled individuals are related to some extent, even if very

distantly, because the population size is finite. In human populations, this sampling scheme corresponds to sampling individuals who are conventionally unrelated. As for the case of realized relationships within families, the GRM is not fixed. We approximate $E(a)$ and $E(b^2)$ in Appendix B. The resulting sampling variance of the estimate of heritability is

$$\text{var}(\hat{h}^2) \approx 2/[E(a) - E(b^2)]/N = 2/[N^2 v(\theta)], \quad (8)$$

where $v(\theta)$ is the variance of relatedness in the population, which is a function of effective population size (Goddard 2009; Goddard *et al.* 2011). Analogous to the within-family design, the sampling variance is inversely proportional to the square of the sample size, rather than by $1/N$ in pedigree designs. Rijdsdijk and Sham (2002) derived the same result (parameterized as the noncentrality-parameter, NCP, of the test statistic for heritability) for QTL linkage mapping in pedigrees, assuming that the variance in relatedness is small. Equation 8 was previously derived for SNP-based estimation of variance components from linear regression theory, assuming that the phenotypic variance is known without error (Vinkhuyzen *et al.* 2013; Visscher *et al.* 2014).

Statistical Power

The interest in this study is not about hypothesis testing but about quantifying the sampling variance of the estimate of heritability. For a detailed treatment on statistical power in variance component estimation using (restricted) maximum likelihood we refer to previous publications (Self and Liang 1987; Shaw 1987; Thompson and Shaw 1990; Almasy and Blangero 1998; Williams and Blangero 1999; Rijdsdijk *et al.* 2001; Purcell *et al.* 2003; Raffa and Thompson 2014). Here we briefly consider the expected value of two test statistics that have been used for hypothesis testing in variance component estimation, the Wald test, and the likelihood-ratio-test statistic.

The Wald test is based on $\hat{h}^4/\text{var}(\hat{h}^2)$, which under the null hypothesis that $h^2 = 0$, follows a χ^2 distribution. However, if $h^2 > 0$, the Wald test statistic follows approximately a noncentral χ^2 with noncentrality parameter (NCP_W)

$$\begin{aligned} \text{NCP}_W &= h^4/\text{var}(\hat{h}^2) = \frac{1}{2}h^4(a - b^2/N) \\ &= \frac{1}{2}h^4 \left[\left(\sum [(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2] \right) \right. \\ &\quad \left. - \left(\sum [(\lambda_i - 1) / (1 + h^2(\lambda_i - 1))] \right)^2 / N \right]. \end{aligned} \quad (9)$$

If the estimation of phenotypic variance is ignored, then

$$\text{NCP}_W = \frac{1}{2}h^4 a = \frac{1}{2}h^4 \left[\sum [(\lambda_i - 1)^2 / (1 + h^2(\lambda_i - 1))^2] \right]. \quad (10)$$

Alternatively, the null hypothesis that $h^2 = 0$ can be tested with a likelihood-ratio test. Blangero and colleagues (Blangero *et al.* 2013) presented a very simple equation for the ELRT statistic to test the null hypothesis of $h^2 = 0$,

$$\text{NCP}_{\text{LRT}} = - \sum \ln[1 + h^2(\lambda_i - 1)]. \quad (11)$$

Equation 11 converges to Equation 10 when $h^2(\lambda_i - 1) \rightarrow 0$. For pairs of relatives with relationship r , $\text{NCP}_W = \frac{1}{2}Nr^2h^4 / (1 - r^2h^4)$ and $\text{NCP}_{\text{LRT}} = -\frac{1}{2}N \ln(1 - r^2h^4)$. These expressions are equivalent when $r^2h^4 \rightarrow 0$. When the true parameter is far from the one being tested under the null, these expressions can give quite different values. Raffa and Thompson (2014) give an analysis based on asymmetrical confidence intervals for the heritability.

Numerical Examples

Figure 1 shows the approximation to the standard error of an estimate of heritability as a function of the population value, experimental sample size, and design. Four different designs were used: a pedigree design of unrelated full-sib pairs, a pedigree design with MZ and DZ twins pairs with a ratio of 1:2 MZ and DZ pairs, a within-family design using full-sib pairs, and a population design using nominally unrelated individuals. In the last two designs, GRM are estimated with SNP data. These designs are less powerful than the pedigree-based experimental designs, but make fewer assumptions. At $N = 10,000$ the sampling variance of the population design approaches that of the pedigree designs, and at $N = 100,000$ it becomes the most powerful design. Sample sizes of 100,000 are realistic in human population and even larger samples sizes are expected in the next few years. Therefore, strong inference on heritability can be drawn using random samples from the population, while not having to make assumptions about the resemblance between relatives due to common environmental factors. The within-family design, which is the most robust with respect to assumptions of the model, remains inaccurate even when the analysis is on 50,000 full-sib pairs. However, in species such as fish with huge full-sib family sizes, accurate estimation could be achieved (Odegard and Meuwissen 2012; Hill 2013).

Figure 2 shows results for the population design for species with different N_e values of 1000, 10,000, and 100,000. It shows the increase in sampling variation with increasing effective population size, which is due to the decrease in the variation in relatedness. For the within-family design the sampling variance of heritability does not depend on the effective population size.

Discussion

We have presented a general framework to quantify the sampling variance of heritability as a function of its population value, the sample size (N), and experimental design.

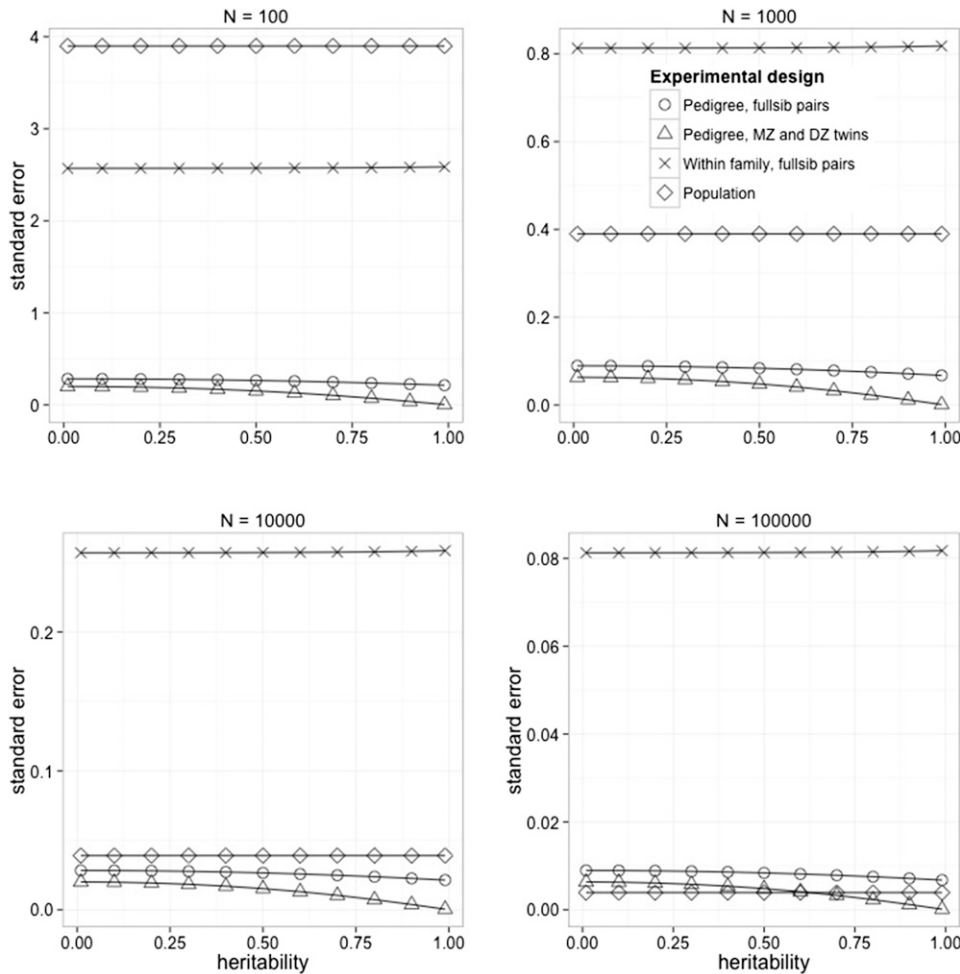


Figure 1 Standard error of estimates of heritability from different experimental designs in human populations, as a function of the population value of the heritability (x -axis), experimental sample size, and experimental design. For the within-family design (*Within-family estimation using realized relationships estimates from markers*), the variance in realized relationships was assumed to be 0.039^2 . For the population design (*Random sampling from the population*), the variance is relatedness was approximated assuming $N_e = 10,000$, a genome length of 35 M, and an average chromosome length of 1 M (Goddard 2009).

Figure 1 shows that the sampling variance is relatively insensitive to the true value of h^2 except when $h^2 \rightarrow 1$. The results recapitulate results from balanced designs and show that for pedigree designs, the sampling variance tends to be proportional to $1/N$. In contrast, for designs that use genetic markers to estimate relatedness within families or estimate relatedness among randomly sampled individuals, the sampling variance is proportional to $1/N^2$. Consequently, very large samples of “unrelated” individuals are powerful for estimating h^2 . The key feature of the experimental design is the variation in relatedness. This is small within families of full-sibs and consequently the sampling variance of h^2 is large.

There are a number of limitations to our study. First, we have assumed that the parameter whose sampling variance we derive is the same in different experimental designs. Even in the absence of confounding factors such as common environmental effect or nonadditive genetic factors, this is not necessarily the case. For the pedigree and within-family design, the parameter given our model assumptions is the narrow-sense heritability. But for the population design it is the proportion of phenotypic variance captured by genetic markers. If these markers are not sufficiently correlated with the genetic variants that cumulatively contribute to the total

narrow sense heritability, then the use of a marker-based GRM will estimate additive genetic variation that is less than the total additive genetic variance. This can occur if the properties of the markers used to create the GRM are different from the segregating causal variants, for example, if the GRM is based upon common SNPs and the causal variants have lower heterozygosity, leading to loss of information due to imperfect linkage disequilibrium (Yang *et al.* 2010). Although a “marker heritability” is conditional on the markers used to estimate relatedness, it is a valid population parameter with predictable sampling properties (as shown in this study). In human populations, it has been used to address the question of “missing heritability” from genome-wide association studies (Yang *et al.* 2010).

Second, we assume that all resemblance between relatives is due to additive genetic covariance, so that there are only two random effects in the model. Additional random effects, for example, common environmental effects, make the covariance matrix \mathbf{V} more complicated and generally not diagonalizable. When there are additional variance components, the residual variance as used in this study is partitioned in two or more components. These additional components are also estimated with error and will have a sampling covariance with the estimate of heritability. We suspect that

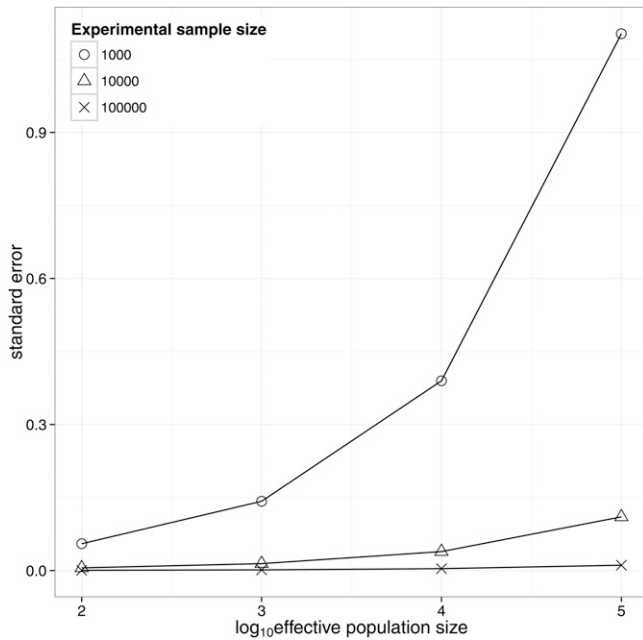


Figure 2 Standard error of the estimate of heritability from random samples of individuals from populations with different effective size and SNP-derived relationship matrices. For each population, a genome length of 35 M and an average chromosome length of 1 M was assumed (Goddard 2009).

having additional variance components in the model will tend to increase the sampling variance of the heritability, except for some balanced designs. However, we have not investigated general properties for designs with multiple random effects. With more than two variance components, computer simulation might be an efficient way to quantify the sampling variance of heritability and the proportion of variance due to additional random effects.

A third assumption is that estimation is by maximum likelihood or, alternatively, that fixed effects and covariates have been adjusted for without error. In practice, researchers tend to use least squares for balanced designs and restricted maximum likelihood (REML) or Bayesian methods for unbalanced designs. The difference in sampling variance between ML and REML is small when there are few fixed effects relative to the sample size, as, for example, in human genetic applications, but larger in situations where there are many fixed effects (e.g., in livestock applications).

Recently, Raffa and Thompson (2014) extended the work of Blangero *et al.* (2013) by deriving approximations to the ELRT and confidence intervals of the heritability estimate using Taylor series expansions of the expected likelihood-ratio test with respect to the distribution of the eigenvalues of a given pedigree. Their simplest approximation can be expressed as an approximate sampling variance of the estimate of heritability as $2/[(N - 1)\text{var}(\lambda)] \approx 2/(N \text{var}(\lambda))$. This expression is the same as our special cases $h^2 \rightarrow 0$ and $\text{All } \lambda_i \rightarrow 1$. The authors show that this approximation is not accurate when the assumptions break down, in particular when eigenvalues are not closely distributed around the

mean of 1, and provide a better approximation using the logarithm of the eigenvalues (Raffa and Thompson 2014). They also show that confidence intervals of the estimates of heritability are not symmetrical when the variance in eigenvalues is large and that Wald statistic-based confidence intervals can be too narrow, implying that the use of the derived standard errors in our study to construct a confidence interval can be anticonservative. Although the derivations from Raffa and Thompson were for a pedigree design, they should also apply to other experimental designs, such as those where GRMs are estimated from marker data.

In conclusion, we have proposed a general unified framework to assess the sampling variance of the estimate of heritability using pedigree or marker-based relationships and have quantified how the sampling variance depends on sample size and the variation in relatedness.

Acknowledgments

This study was inspired by John Blangero's presentation at the 2013 Statistical and Quantitative Genetics conference in Seattle. We thank Bill Hill for discussions and helpful comments, Jesse Raffa for useful comments, and Matt Robinson and Kostya Shakhbazov for feedback and help with R. This research was supported by U.S. National Institutes of Health (NIH) grant R01 GM075091.

Literature Cited

- Almasy, L., and J. Blangero, 1998 Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* 62: 1198–1211.
- Berenos, C., P. A. Ellis, J. G. Pilkington, and J. M. Pemberton, 2014 Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Mol. Ecol.* 23: 3434–3451
- Blangero, J., V. P. Diego, T. D. Dyer, M. Almeida, J. Peralta *et al.*, 2013 A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv. Genet.* 81 (81): 1–31.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Longman, Harlow, Essex, United Kingdom.
- Goddard, M., 2009 Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136: 245–257.
- Goddard, M. E., B. J. Hayes, and T. H. Meuwissen, 2011 Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128: 409–421.
- Hemani, G., J. Yang, A. Vinkhuyzen, J. E. Powell, G. Willemsen *et al.*, 2013 Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am. J. Hum. Genet.* 93: 865–875.
- Hill, W. G., 2013 On estimation of genetic variance within families using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* 45: 32.
- 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.
- Janss, L., G. de Los Campos, N. Sheehan, and D. Sorensen, 2012 Inferences from genomic models in stratified populations. *Genetics* 192: 693–704.

- Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson *et al.*, 2011 FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8: 833–835.
- Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- Neale, M., and L. R. Cardon, 1992 *Methodology for Genetic Studies of Twins and Families*. Kluwer, Dordrecht, The Netherlands.
- Odegard, J., and T. H. Meuwissen, 2012 Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* 44: 16.
- Patterson, H. D., and R. Thompson, 1971 Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545–554.
- Purcell, S., S. S. Cherny, and P. C. Sham, 2003 Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19: 149–150.
- Raffa, J. D., and E. A. Thompson, 2014 Power and effective study size based on approximations to the expected likelihood ratio test in heritability studies. Department of Statistics, University of Washington, Technical Report #630. Available at: <http://www.stat.washington.edu/research/reports/2014/tr630.pdf>
- Rijsdijk, F. V., and P. C. Sham, 2002 Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* 3: 119–133.
- Rijsdijk, F. V., J. K. Hewitt, and P. C. Sham, 2001 Analytic power calculation for QTL linkage analysis of small pedigrees. *Eur. J. Hum. Genet.* 9: 335–340.
- Ritland, K., 2000 Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* 9: 1195–1204.
- Robinson, M. R., A. W. Santure, I. Decauwer, B. C. Sheldon, and J. Slate, 2013 Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Mol. Ecol.* 22: 3963–3980.
- Searle, S. R., 1982 *Matrix Algebra Useful for Statistics*. Wiley, New York.
- Self, S. G., and K. Y. Liang, 1987 Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* 82: 605–610.
- Shaw, R. G., 1987 Maximum-likelihood approaches applied to quantitative genetics of natural-populations. *Evolution* 41: 812–826.
- Stranden, I., and D. J. Garrick, 2009 Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92: 2971–2975.
- Thomas, S. C., 2005 The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1457–1467.
- Thompson, E. A., and R. G. Shaw, 1990 Pedigree analysis for quantitative traits: variance-components without matrix-inversion. *Biometrics* 46: 399–413.
- Thompson, E. A., and R. G. Shaw, 1992 Estimating polygenic models for multivariate data on large pedigrees. *Genetics* 131: 971–978.
- Thompson, R., and K. D. Atkins, 1994 Sources of information for estimating heritability from selection experiments. *Genet. Res.* 63: 49–55.
- VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414–4423.
- Vinkhuyzen, A. A., N. R. Wray, J. Yang, M. E. Goddard, and P. M. Visscher, 2013 Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu. Rev. Genet.* 47: 75–95.
- Visscher, P. M., S. E. Medland, M. A. Ferreira, K. I. Morley, G. Zhu *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2: e41.
- Visscher, P. M., S. Macgregor, B. Benyamin, G. Zhu, S. Gordon *et al.*, 2007 Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* 81: 1104–1110.
- Visscher, P. M., G. Hemani, A. A. Vinkhuyzen, G. B. Chen, S. H. Lee *et al.*, 2014 Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 10: e1004269.
- Williams, J. T., and J. Blangero, 1999 Power of variance component linkage analysis to detect quantitative trait loci. *Ann. Hum. Genet.* 63: 545–563.
- 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.

Communicating editor: L. E. B. Kruuk

Appendix A: Derivation of the Sampling Variance of Heritability from Within-Family Designs

As before, $\mathbf{y} = \mathbf{g} + \mathbf{e}$, with $\text{var}(\mathbf{g}) = \mathbf{G}\sigma_g^2$ and $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$. If all individuals belong to a family, $E(\mathbf{G}) = \mathbf{I}$ on diagonals and θ on off-diagonals. We extract the family mean (u) from an individual's breeding value so $\mathbf{y} = u + \mathbf{g}^* + \mathbf{e}$, where $\text{var}(\mathbf{g}^*) = (\mathbf{G} - \theta\mathbf{J})\sigma_g^2 = \mathbf{W}\sigma_g^2$. If we treat u as fixed, then $\text{var}(\mathbf{y} - u) = \mathbf{W}\sigma_g^2 + \mathbf{I}\sigma_e^2$. The mean eigenvalue of \mathbf{W} is $(1 - \theta)$ and the variance of eigenvalues is $n \times \text{var}(r_{ij})$ where r_{ij} is the realized relationship between individuals i and j within the same family.

As in Hill(2013), we derive the sampling variance for a single family of size n . Under our assumed model of no environmental effects shared by family members, $t = \theta h^2$. As before, the elements of the information matrix are $\mathbf{F}_{11} = 1/2a$, $\mathbf{F}_{12} = \mathbf{F}_{21} = \frac{1}{2}b/\sigma^2$, $\mathbf{F}_{22} = \frac{1}{2}n/\sigma^4$. We approximate the elements of the information matrix by taking a second-order Taylor series about the mean eigenvalue of $(1 - \theta)$. Then, approximately,

$$E(a)/n = \theta^2 / (1-t)^2 + (1 + 2\theta h^2)n \times \text{var}(r_{ij}) / (1-t)^4$$

$$E(b)/n = -\theta / (1-t) - h^2 n \times \text{var}(r_{ij}) / (1-t)^3.$$

Using these to construct the $(\mathbf{F}^{-1})_{1,1}$ gives

$$\text{var}(h^2) \approx \left[2(1-t)^2 / (n^2 \times \text{var}(r_{ij})) \right] \left[(1-t)^2 - nh^4 \text{var}(r_{ij}) \right].$$

Using this approximation, the determinant of the information matrix, and therefore the approximation of the sampling variance of heritability, can be negative, when $n > (1-t)^2 / (h^4 \times \text{var}(r_{ij}))$. For example, for full-sibs ($\text{var}(r_{ij}) \sim 0.038^2$) and $h^2 = 0.8$ (and $t = 0.4$), a sampling size of $n = 390$ full-sibs would result in a predicted sampling variance of the estimate of heritability that is negative. Presumably a higher-order Taylor series would correct this, but at the expense of having a relatively simple expression.

If we now use the eigenvalue decomposition of \mathbf{W} , as in Thompson and Atkins (1994) and as used in our other designs but parameterizing the variance components instead of h^2 and σ^2 , then the element of the Information matrix (\mathbf{S} in the Hill notation) are

$$S_{11} = \frac{1}{2}a = \frac{1}{2} \sum \left[\lambda_i^2 / (\sigma_e^2 + \lambda_i \sigma_g^2)^2 \right],$$

$$S_{12} = \frac{1}{2}b = \frac{1}{2} \sum \left[\lambda_i / (\sigma_e^2 + \lambda_i \sigma_g^2)^2 \right],$$

$$S_{22} = \frac{1}{2}c = \frac{1}{2} \sum \left[1 / (\sigma_e^2 + \lambda_i \sigma_g^2)^2 \right].$$

If we take expectations of a , b , and c , where the expectation is over the eigenvalues of \mathbf{W} [with mean = $(1 - \theta)$, variance = $\text{var}(\lambda)n \times \text{var}(r_{ij})$], then, from a second-order Taylor series about the mean:

$$E(a)/n = (1-\theta)^2 / (1-t)^2 + \left[\text{var}(\lambda) / (1-t)^4 \right] \left[(1-h^2)^2 - 2(1-\theta)h^2(1-h^2) \right],$$

$$E(b)/n = (1-\theta) / (1-t)^2 + \left[\text{var}(\lambda) / (1-t)^4 \right] \left[(1-\theta)h^4 - 2h^2(1-h^2) \right],$$

$$E(c)/n = 1 / (1-t)^2 + \left[\text{var}(\lambda) / (1-t)^4 \right] \left[3h^4 \right].$$

Finally, the sampling variance of the estimate of σ_g^2 is, approximately,

$$\begin{aligned} v(\sigma_g^2) &\approx \left[2(1-t)^2 / (n \times \text{var}(\lambda)) \right] \left[(1-t)^2 + 3h^4 \text{var}(\lambda) \right] / \left[(1-t)^2 + -h^4 \text{var}(\lambda) \right] \\ &= \left[2(1-t)^2 / n^2 \text{var}(r_{ij}) \right] \left[(1-t)^2 + 3nh^4 \text{var}(r_{ij}) \right] / \left[(1-t)^2 - nh^4 \text{var}(r_{ij}) \right]. \end{aligned}$$

These terms are similar but not identical to Hill(2013). The difference is because we use ML whereas Hill used REML and we have assumed that the family mean is fixed. For very large n the above expression converges to that given by Hill(2013).

Appendix B: Derivation of $E(a)$ and $E(b)$ for Population Designs

Let $x_i = \lambda_i - 1$, so that $a = \sum x_i^2 / (1 + x_i h^2)$, and $b = \sum x_i / (1 + x_i h^2)$. A second-order Taylor series expansion around $x \rightarrow 0$ gives $E(a) = N \text{var}(\lambda)$ and $E(b^2) = E(a)$.

The variance of eigenvalues is derived from the GRM (\mathbf{G})

$$\mathbf{G} = \mathbf{T}\mathbf{D}\mathbf{T}'$$

with diagonal matrix \mathbf{D} containing the eigenvalues λ . \mathbf{G} can also be written as

$$\mathbf{G} = \mathbf{I} + \mathbf{\Delta},$$

with \mathbf{I} the identify matrix and $\mathbf{\Delta}$ a matrix containing small relationships between distantly related individuals. Element Δ_{ij} are random with $E(\Delta_{ij}) = 0$ and $\text{var}(\Delta_{ij}) = \nu(\theta)$, the variance in relatedness in the population. $\nu(\theta)$ is a function of effective population size (Goddard 2009; Goddard *et al.* 2011),

$$\mathbf{G}^2 = \mathbf{T}\mathbf{D}^2\mathbf{T}'$$

$$\text{tr}(\mathbf{G}^2) = \sum \lambda_i^2 = N[1 + \text{var}(\lambda)], \text{ since } E(\lambda) = 1$$

$$\mathbf{G}^2 = [\mathbf{I} + \mathbf{\Delta}]^2 = \mathbf{I} + 2\mathbf{\Delta} + \mathbf{\Delta}^2,$$

with

$$\text{tr}(\mathbf{G}^2) = N + 0 + \text{tr}(\mathbf{\Delta}^2) = N + N^2 \nu(\theta),$$

since $\text{tr}(\mathbf{\Delta}^2)$ is the sum of squares of element in $\mathbf{\Delta}$, each with expectation $\nu(\theta)$.

Hence, we have $E(a) = E(b^2) = N \text{var}(\lambda) = N^2 \nu(\theta)$. Therefore, $\text{var}(\lambda) = N\nu(\theta)$ and proportional to experimental sample size. Finally,

$$\text{var}(h^2) \approx 2/[E(a) - E(b^2)/N] = 2/[E(a)(1 - 1/N)] \approx 2/E(a) = 2/[N^2 \nu(\theta)].$$

The variance in relatedness is $\nu(\theta) = \sum \sum r_{ij}^2$, the sum of linkage disequilibrium correlations r^2 over all pairs of SNPs that are used to construct the GRM (Goddard 2009; Goddard *et al.* 2011).