

On the Sampling Variance of Intraclass Correlations and Genetic Correlations

Peter M. Visscher

University of Edinburgh, Institute of Ecology and Resource Management, Edinburgh EH9 3JG, Scotland

Manuscript received February 3, 1997

Accepted for publication March 23, 1998

ABSTRACT

Widely used standard expressions for the sampling variance of intraclass correlations and genetic correlation coefficients were reviewed for small and large sample sizes. For the sampling variance of the intraclass correlation, it was shown by simulation that the commonly used expression, derived using a first-order Taylor series performs better than alternative expressions found in the literature, when the between-sire degrees of freedom were small. The expressions for the sampling variance of the genetic correlation are significantly biased for small sample sizes, in particular when the population values, or their estimates, are close to zero. It was shown, both analytically and by simulation, that this is because the estimate of the sampling variance becomes very large in these cases due to very small values of the denominator of the expressions. It was concluded, therefore, that for small samples, estimates of the heritabilities and genetic correlations should not be used in the expressions for the sampling variance of the genetic correlation. It was shown analytically that in cases where the population values of the heritabilities are known, using the estimated heritabilities rather than their true values to estimate the genetic correlation results in a lower sampling variance for the genetic correlation. Therefore, for large samples, estimates of heritabilities, and not their true values, should be used.

THERE are three classic papers on the topic of sampling variances of estimates of genetic correlations in the 1950s: Reeve (1955), who derived expressions of the sampling variance for parent-offspring designs; Robertson (1959a), who derived general expressions for balanced one-way ANOVA designs with equal population values for heritabilities; and Tallis (1959), who derived general expressions for balanced and unbalanced one-way designs. Although the latter article is, in a sense, the most general, it is not referred to frequently (it does not help that the reference added in the proof of the Robertson article points to the wrong journal).

It could be argued that the expressions derived in those articles are no longer relevant, since estimation techniques have moved on from least-squares methods to likelihood-based methods [mainly *residual maximum likelihood* (REML); Patterson and Thompson (1971)]. Using likelihood methods, sampling variances can be approximated from likelihood profiles (Meyer and Hill 1991). However, numerous publications, particularly in the evolutionary genetics literature, still use the expressions derived by Reeve, Robertson, and Tallis. There appears to be some confusion about the use of expressions for the sampling variance of genetic correlations. For example, Koots and Gibson (1996), who performed a meta-analysis of an impressive number (1500) of estimates of heritabilities and genetic correla-

tions for beef cattle traits, argued that using the estimates of heritabilities and genetic correlations in the expressions of Reeve (1955) and Robertson (1959a), rather than their true (population) values, appeared to be closer to the observed empirical sampling variance of the genetic correlation coefficient. This was corroborated by a small simulation study. Furthermore, these authors implied that the equations derived in Reeve (1955) and Robertson (1959a) were inaccurate, because the empirical sampling variance of estimated genetic correlation coefficients was much larger than the expected sampling variance, using both their data and simulations.

Three areas of confusion can be identified:

1. Should the parameters in the expressions for the sampling variance of the genetic correlation coefficient in Reeve (1955), Robertson (1959a), and Tallis (1959) be the population parameters or the estimates of those parameters?
2. How good are the expressions derived in the trio of articles?
3. What is the impact of estimates that are outside the parameter space, *i.e.*, negative heritability estimates and/or estimates of genetic correlations < -1 or $> +1$?

In this article I review the main expressions of Reeve, Robertson, and Tallis and clarify under what circumstances the expressions should be used. I also review and evaluate the equations for the sampling variance of intraclass correlations as a paradigm, since these are

Address for correspondence: University of Edinburgh, Institute of Ecology and Resource Management, West Mains Rd., Edinburgh EH9 3JG, Scotland. E-mail: peter.visscher@ed.ac.uk

central to the understanding of the assumptions and methods.

MATERIALS AND METHODS

To answer the above questions, we first look at the derivation of the sampling variance of intraclass correlation coefficients because this serves as an appropriate paradigm for the sampling variance of the genetic correlation coefficient. Subsequently, expressions for the sampling variance of the genetic correlation are reviewed and evaluated.

Sampling variance of intraclass correlation coefficients

Population parameters known: Consider a simple, balanced one-way design with s sires and n progeny per sire, and assume that parameters are estimated using least-squares methods, *e.g.*, ANOVA. Observations are assumed to be normally distributed. The expectations and variances of the between- and within-sire mean squares (MS) are

$$\begin{aligned} E[B] &= [(1 - \hat{t}) + nt]\sigma_p^2, \\ E[W] &= [(1 - \hat{t})]\sigma_p^2, \\ \text{var}(B) &= 2\{[(1 - \hat{t}) + nt]\sigma_p^2/[s - 1], \\ \text{var}(W) &= 2\{[(1 - \hat{t})]\sigma_p^2/[s(n - 1)]. \end{aligned}$$

with B and W the between- and within-sire MS, t the population intraclass correlation, and σ_p the phenotypic standard deviation. The variance of the least-squares estimate of the intra-class correlation,

$$\begin{aligned} \hat{t} &= [(B - W)/n]/[(B - W)/n + W] \\ &= [B - W]/[B + (n - 1)W] = A/C, \end{aligned}$$

can be approximated using a first-order Taylor-series expansion about the expected mean squares,

$$\begin{aligned} \text{var}(\hat{t}) &\approx \text{var}(A)/E^2(C) + E^2(A)\text{var}(C)/E^4(C) \\ &\quad - 2E(A)\text{cov}(A, C)/E^3(C) \\ &= 2(1 - \hat{t})^2[1 + (n - 1)\hat{t}]^2 [(sn - 1)/ \\ &\quad (n^2s(s - 1)(n - 1))] \\ &\approx 2(1 - \hat{t})^2[1 + (n - 1)\hat{t}]^2/[(s - 1)n(n - 1)]. \end{aligned} \quad (1)$$

Equation 1 is well known (Robertson 1959b; Falconer and Mackay 1996). Its derivation, using means and variances of mean squares, appears to be first given by Osborne and Paterson (1952).

For large n , Equation 1 reduces to

$$\text{var}(\hat{t}) \approx 2t^2 (1 - \hat{t})^2/(s - 1). \quad (2)$$

A number of expressions similar to Equation 1, which all differ in the terms relating to the number of sires and progeny per sire, can be found in the literature. They differ in particular in the degrees of freedom relating to the between-sire component of variance. The general expression is

$$\text{var}(\hat{t}) \approx f(\hat{t}) c_1 \quad (3)$$

with

$$f(\hat{t}) = 2(1 - \hat{t})^2[1 + (n - 1)\hat{t}]^2,$$

Where c_1 is a function of s and n , from reference i . The original expression was derived in a classical paper by Fisher (1921),

$$c_1 = 1/[sn(n - 1)]. \quad (4)$$

Robertson and Lerner (1949) quoted Fisher, but used

$$c_2 = 1/[(s - 2)n(n - 1)]. \quad (5)$$

Finally, following Zerbe and Goldgar (1980), an expression can be derived using

$$\begin{aligned} y &= B/W = [1 + (n - 1)\hat{t}]/[1 - \hat{t}] \\ &\sim \{[1 - \hat{t}]/[1 + (n - 1)\hat{t}]\} F_{s-1, s(n-1)}. \end{aligned}$$

Hence,

$$\text{var}(\hat{t}) \approx [d\hat{t}/dy]^2 \text{var}(y) = f(\hat{t}) c_3$$

with

$$\begin{aligned} c_3 &= [s(n - 1)]^2 [sn - 3]/[n^2(s - 1) \\ &\quad (s(n - 1) - 2)^2 (s(n - 1) - 4)]. \end{aligned} \quad (6)$$

Although this expression is trivial to derive, to my knowledge this is the first time that it has been documented following a first-order approximation using the F ratio.

All the above expressions for c_i reduce to the one most commonly used (*i.e.*, $c_0 = 1/[n(s - 1)(n - 1)]$) for large s and n . However, even for large n , there are discrepancies between the formulas depending on the use of s , $(s - 1)$, or $(s - 2)$ in the denominator. It is likely that the form used by Robertson and Lerner is incorrect and probably stems from the use of Fisher's z -transformation. Fisher showed that

$$\text{var}(z) \approx n/[2(n - 1)(s - 2)],$$

with

$$z = \frac{1}{2} \log\{[1 + (n - 1)\hat{t}]/[1 - \hat{t}]\}.$$

Hence,

$$\begin{aligned} \text{var}(\hat{t}) &\approx [d\hat{t}/dz]^2 \text{var}(z) = \{4(1 - \hat{t})^2[1 + (n - 1)\hat{t}]^2/n^2\} \\ &\quad \{n/[2(n - 1)(s - 2)]\} \\ &= f(\hat{t}) c_2. \end{aligned}$$

However, as pointed out by Osborne and Paterson (1952), this is a roundabout way of deriving the sampling variance of the estimated intraclass correlation, by first transforming to z and then back again.

Population values unknown: Except when doing power calculations and/or investigating the design of experiments (Robertson 1959b) or simulation studies, we do not know the population values and hence do not know the exact or approximate sampling variance of the intraclass correlation. The standard practice is to use the formulae derived in the previous section and to substitute \hat{t} for t . This is essentially based upon the assumption that $E(\hat{t}) = t$. However, it should be obvious that if, by chance, the estimate of the intraclass correlation is too high (or too low), the resulting estimate of the sampling variance will be biased. Using

$$\begin{aligned} d\{f(\hat{t})\}/d\hat{t} &= 4[1 - \hat{t}][1 + (n - 1)\hat{t}] \\ &\quad [(n - 1)(1 - 2\hat{t}) - 1] \end{aligned} \quad (7)$$

gives a maximum estimate of the sampling variance of $\hat{t} = \frac{1}{2} [(n - 2)/(n - 1)]$ (see also Taylor 1976). The minimum estimate of the sampling variance is found for $\hat{t} = 1$ or $\hat{t} = -1/(n - 1)$. Only when there is a scale on which the sampling variance is (nearly) independent of the population values, for example, Fisher's z -scale, will the estimate of the sampling variance be correct.

Kemthorne (1957) argued that an adjustment to the degrees of freedom should be made when estimating the sam-

pling variance of an estimate of the intraclass correlation, since

$$\begin{aligned}\text{var}(\text{MS}) &= 2E^2(\text{MS})/\text{d.f.}, \\ E[2\text{MS}^2/(\text{d.f.} + 2)] &= 2[\text{var}(\text{MS}) + E^2(\text{MS})]/(\text{d.f.} + 2) \\ &= 2[E^2(\text{MS})\{2/\text{d.f.} + 1\}]/(\text{d.f.} + 2) \\ &= 2E^2(\text{MS})/\text{d.f.} = \text{var}(\text{MS}),\end{aligned}$$

where d.f. is degrees of freedom. This indicates an unbiased estimate of the sampling variance as

$$\begin{aligned}\text{var}(\hat{t}) &\approx 2(1 - \hat{t})^2 (1 + (n - 1)\hat{t})^2 \frac{(sn - 3)}{n^2(s + 1)(s(n - 1) + 2)} \\ &= f(\hat{t})c_1.\end{aligned}\quad (8)$$

Equation 8 reduces to the standard equation of Osborne and Paterson (1952) and Robertson (1959b) for large sn .

Simulation study: Simulations were performed to compare the empirical standard deviation of heritability estimates, the predicted standard deviation using the true population values (Equation 1), and the average estimated standard deviation (using Equations 1, 4, and 8, with \hat{t} substituted for t). Independent between-sire and within-sire sums of squares were sampled from central χ^2 distributions with $(s - 1)$ and $[s(n - 1)]$ degrees of freedom, respectively, and then scaled to the appropriate mean squares using the population values of t and σ_p and the values of s and n . Without loss of generality, a phenotypic standard deviation of unity was used throughout.

Since the only difference between the various prediction equations for the sampling variances are functions of s and n , only results for the Taylor series (Osborne and Paterson 1952) are presented. The other equations for the sampling variance differ approximately by factors of $(s - 1)/s$ (using Fisher's formula) and $(s - 1)/(s + 1)$ (using Kempthorne 1957).

Sampling variance of genetic correlation coefficient

Expressions from literature: Tallis (1959) derived a general expression for the approximate sampling variance of the estimated genetic correlation coefficient for a balanced half-sib design. Population values for the intraclass correlations of the two traits are t_1 and t_2 , and the genetic and within-sire correlation coefficients are r_g and r_w , respectively. The general form of Tallis' expression is

$$\text{var}(\hat{r}_g) \approx P/(s - 1) + Q/(s(n - 1)),$$

for

$$\begin{aligned}P &= (1 + r_g^2)[1/(R_1 R_2) + A^2] - 2r_g(1/R_1 + 1/R_2)A \\ &\quad + r_g^2(t_1 - t_2)^2/[2(nt_1 t_2)^2]\end{aligned}\quad (9)$$

and

$$\begin{aligned}Q &= (1/R_1 - 1)(1/R_2 - 1)[(1 + r_g^2)(1 + r_w^2) \\ &\quad - 2r_g(A - r_g)] + r_g^2(t_1 - t_2)^2/[2(nt_1 t_2)^2]\end{aligned}\quad (10)$$

with

$$A = r_g + r_w[(1/R_1 - 1)(1/R_2 - 1)]^{1/2}.\quad (11)$$

$R_i = n/(n + (1 - t_i)/t_i)$, which is the general expression of the reliability of a progeny test based upon n progeny and a heritability of $4t_i$. Tallis (1959) used a different (but equivalent) expression for P and Q , but we find it more convenient to write the terms as presented here.

Special cases:

$$1. \quad t_1 = t_2 = t.$$

This is the scenario of Robertson (1959a). The terms P and Q simplify to

$$\begin{aligned}P(t_1 = t_2) &= (1/R^2)[(1 + r_g^2)(1 + \{r_g R + r_w(1 - R)\}^2) \\ &\quad - 4r_g(r_g R + r_w(1 - R))],\end{aligned}\quad (12)$$

$$Q(t_1 = t_2) = [(1 - R)/R]^2 [(1 + r_g^2)(1 + r_w^2) - 4r_g r_w]\quad (13)$$

with

$$R = n/(n + (1 - t)/t).$$

These correspond to the equations given by Robertson (1959a, p. 473), although his P and Q were scaled by a factor of $(nt)^2$.

$$2. \quad t_1 = t_2 = t \quad \text{and} \quad r_g = r_w = r.$$

A further simplification is if the genetic and within-sire correlation are the same. The expression for the sampling variance may be written as

$$\begin{aligned}\text{var}(\hat{r}_g) &\approx [(1 - r^2)^2/R^2][1/(s - 1) \\ &\quad + (1 - R)^2/(s(n - 1))],\end{aligned}\quad (14)$$

which corresponds to Robertson's formula on page 474. In general, when $t_1 \neq t_2$, there is no simple form for the sampling variance of the genetic correlation coefficient.

Finally, Robertson (1959a) suggested a very simple and general expression for the sampling variance of the genetic correlation coefficient by observing the similarity between expressions derived for special cases,

$$\text{var}(\hat{r}_g) \approx (1 - r_g^2)^2 [\text{var}(\hat{h}_1^2)\text{var}(\hat{h}_2^2)]^{1/2} / [2h_1^2 h_2^2].\quad (15)$$

This was the equation used by Koots and Gibson (1996).

$$3. \quad n \rightarrow \infty, \quad \text{i.e., } R_i = 1.$$

For a large number of progeny per sire, Tallis' equation reduces to a very simple form,

$$\text{var}(\hat{r}_g) \approx (1 - r_g^2)^2/(s - 1).\quad (16)$$

This equation is equivalent to the approximation of the sampling variance of a correlation coefficient in the bivariate normal case with $(s - 1)$ degrees of freedom.

There are difficulties in using the approaches for the sampling variance of the intraclass correlation to determine the sampling variance of estimates of the genetic correlation coefficient: (1) The estimate of r_g is unbounded in principle, so that large positive and negative values (outside the range -1 to $+1$) are possible, and (2) the true heritability, or its estimate, appears in both the numerator and the denominator of the equation for the sampling variance of r_g (see Equations 9 to 11). This means that in the vicinity of true or estimated h^2 being zero, the estimate of the sampling variance can become very large because of a division by a small number. Also, if one or both of the estimated heritabilities is < 0 , the estimate of the genetic correlation coefficient is an imaginary number (Reeve 1955). Simulation results are less meaningful in these cases, because the estimate of the sampling variance may not have converged (or never will). For example, in the simulation study of parent-offspring regression in Koots and Gibson (1996), with true population parameters of $h_1^2 = h_2^2 = 0.10$ and $r_g = 0$, the estimates of r_g varied from -2.5 and $+4.18$, and the empirical variance of estimated genetic correlation coefficients was very large when both estimates of the heritability were close to zero.

Heritabilities known: Koots and Gibson (1996) argued that in some (rare) cases, the population heritabilities may be known when the genetic correlation is estimated, for example, through prior information or a meta-analysis of relevant litera-

ture values. In that case, one could proceed with only estimating the between- and within-sire covariances, along with the phenotypic variances, from the data. If the phenotypic variances are assumed to be estimated accurately, *i.e.*, for large sn , then

$$\hat{r}_g \approx [(B_{12} - W_{12})/n]/[t_1 t_2 \sigma_{p1}^2 \sigma_{p2}^2]^{1/2},$$

with B_{12} and W_{12} the between-sire mean crossproduct and within-sire mean crossproduct, respectively, and t_i and σ_{pi}^2 the known intraclass correlations and phenotypic variances for trait i . The means and variances of the crossproducts, using W_{ii} and B_{ii} to denote the within and between-sire MS for trait i , are

$$E(W_{12}) = (r_w[(1 - t_1)(1 - t_2)]^{1/2})\sigma_{p1}\sigma_{p2},$$

$$E(B_{12}) = (r_w[(1 - t_1)(1 - t_2)]^{1/2} + r_g(t_1 t_2)^{1/2})\sigma_{p1}\sigma_{p2},$$

$$\text{var}(W_{12}) = [E(W_{11})E(W_{22}) + E(W_{12})^2]/(s(n - 1)),$$

$$\text{var}(B_{12}) = [E(B_{11})E(B_{22}) + E(B_{12})^2]/(s - 1),$$

$$\text{cov}(W_{12}, B_{12}) = 0.$$

Using

$$\begin{aligned} \text{var}(\hat{r}_g) &= [\text{var}(B_{12}) + \text{var}(W_{12})]/(r^2 t_1 t_2 \sigma_{p1}^2 \sigma_{p2}^2) \\ &= P/(s - 1) + Q/(s(n - 1)), \end{aligned} \quad (17)$$

the sampling variance of the genetic correlation coefficient can be calculated by substituting the expressions for $E(B_{12})$, $E(W_{12})$, $\text{var}(B_{12})$, and $\text{var}(W_{12})$ into Equation 17. In particular, for $t_1 = t_2 = t$,

$$P(t_1 = t_2) = (1/R^2)[1 + \{r_g R + (1 - R)r_w\}^2] \quad (18)$$

and

$$Q(t_1 = t_2) = [(1 - R)/R]^2 (1 + r_w^2). \quad (19)$$

The values of P and Q are larger when the heritabilities are assumed known; *i.e.*, the sampling variance of the genetic correlation is larger when heritabilities are assumed known (*cf.* Equations 12 and 13). This is most clearly seen when the number of progeny are very large, because then

$$\text{var}(\hat{r}_g) = (1 + r_g^2)/(s - 1), \quad (20)$$

which is always larger than the derivation for unknown heritabilities (Equation 16). These findings are in agreement with Koots and Gibson (1996), who argued that the estimated heritabilities, and not the population values (if known), should be used to estimate the sampling variance of the genetic correlation coefficient. The results arise because of the positive covariance between the estimated between-sire covariance (from B_{12}) and between-sire variances (from B_{11} and B_{22}). Hence, the absolute value of the genetic correlation coefficient and the estimates of the heritabilities are positively correlated (unless $r_g = 0$), which is taken into account by using the estimated rather than true values of the heritabilities in the expressions to estimate the sampling variance of the genetic correlation.

Simulations: A simulation study was performed in which the empirical variance of the estimated r_g and the average estimated sampling variance using the estimates of heritabilities and correlations was compared to the predicted sampling variance [using equations from Tallis (1959); see above]. Independent 2×2 matrices of within-sire mean squares and mean crossproducts (W_{11} , W_{12} , and W_{22}) and between-sire mean squares and crossproducts (B_{11} , B_{12} , and B_{22}) were sampled from a central Wishart distribution with $(s(n - 1))$ and $(s - 1)$ degrees of freedom, respectively (see Visscher 1995 for more details). The resulting matrices of mean squares and cross-

products were used to calculate least-squares estimates of the two heritabilities and the genetic and within-sire correlation coefficients in the standard way (see, for example, Tallis 1959). For estimated sire variances that were positive, estimates of genetic correlations could be < -1 or > 1 . A least-squares estimate of the genetic correlation coefficient was not possible when one or both of the heritability estimates were negative.

Since most authors (*e.g.*, Fisher 1921; Robertson 1959a; Tallis 1959) have explicitly warned against the use of the "standard" equations when the true heritability, or its estimate, is close to zero, it seems more meaningful to force the estimate of r_g to be > -1 and < 1 . Therefore, if the estimate of the genetic covariance matrix, *i.e.*, $(\mathbf{B} - \mathbf{W})/n$, with \mathbf{B} and \mathbf{W} the 2×2 matrices of between- and within-sire mean squares, respectively, was not positive-definite, REML estimates were calculated using the sampled between- and within-sire mean squares and crossproducts. To force the parameters in the parameter space, a form of "bending" the (least-squares) covariance matrix was applied (*e.g.*, Hayes and Hill 1981; Calvin 1993; Visscher 1995). The form of bending applied was described as attenuating the covariance matrix by Visscher (1995), and estimates of (co)variances are the same if a REML analysis had been carried out on the original data. This was done by calculating heritabilities on a canonical scale and setting negative heritabilities to a small positive value (10^{-6}) and heritabilities larger than one to a value less than one ($1 - 10^{-6}$). Following Koots and Gibson (1996), simulated data sets were summarized only if the geometric mean of the estimated heritabilities was > 0.01 . For each set of parameters, simulation was stopped when 10^5 replicated samples of the estimated genetic correlation coefficient were obtained.

RESULTS

Validation of expressions for sampling variances of intraclass correlations: Predicted sampling variances of heritability estimates, observed sampling variances from simulations, and the average estimated sampling variance from simulation are presented in Table 1. For the predicted and estimated sampling variances, only the equation of Osborne and Paterson (1952) was used. To obtain other predictions, results for the standard error of the heritability need to be multiplied by, approximately, factors of $[(s - 1)/s]^{1/2}$ (using Fisher's formula) and $[(s - 1)/(s + 1)]^{1/2}$ (using Kempthorne 1957). Results (Table 1) indicate that the approximation of Osborne and Paterson (1952) works very well, in that the predicted variation in estimated heritabilities is close to the observed variation from simulation, in particular for a small value of the heritability. For a large heritability and a small number of sires, it appears that the approximation of Fisher is better. For example, for $s = 2$, $n = 1000$, and $h^2 = 0.96$, the predicted standard error from Osborne and Paterson (1952) is 1.0353 (Table 1) and from Fisher's equation, is 0.7323 (not shown in tables), while the observed standard error was 0.7192. However, further simulations (not shown) with different values of the heritability indicated that both predictions deviate substantially from the observed standard errors for large heritabilities and a small number of sires. For example, for $t = 0.95$ (and hence a "heritability" of 3.80), the observed standard error in estimated

TABLE 1
Observed (*O*), estimated (*E*), and predicted (*P*) standard errors of intraclass correlations

| <i>s</i> ^b | <i>nr</i> | <i>h</i> ^{2a} = 0.04 | | | <i>h</i> ² = 0.96 | | |
|-----------------------|-----------|---|---------------------------|---------------------------|---|---------------------------|---------------------------|
| | | <i>O</i> [4σ(<i>h</i>) ^d] | <i>E</i> [4σ(<i>h</i>)] | <i>P</i> [4σ(<i>h</i>)] | <i>O</i> [4σ(<i>h</i>) ^d] | <i>E</i> [4σ(<i>h</i>)] | <i>P</i> [4σ(<i>h</i>)] |
| 2 | 1000 | 0.0593 | 0.0590 | 0.0616 | 0.7192 | 0.6716 | 1.0353 |
| 4 | 1000 | 0.0350 | 0.0352 | 0.0355 | 0.5169 | 0.4996 | 0.5978 |
| 8 | 1000 | 0.0232 | 0.0231 | 0.0233 | 0.3661 | 0.3609 | 0.3914 |
| 16 | 5 | 0.3288 | 0.3269 | 0.3342 | 0.4721 | 0.4648 | 0.4835 |
| | 10 | 0.1648 | 0.1636 | 0.1656 | 0.3575 | 0.3545 | 0.3686 |
| | 50 | 0.0433 | 0.0432 | 0.0435 | 0.2764 | 0.2746 | 0.2860 |
| 32 | 5 | 0.2323 | 0.2308 | 0.2332 | 0.3346 | 0.3309 | 0.3374 |
| | 10 | 0.1148 | 0.1147 | 0.1154 | 0.2537 | 0.2518 | 0.2568 |
| | 50 | 0.0302 | 0.0302 | 0.0303 | 0.1967 | 0.1951 | 0.1990 |
| 100 | 5 | 0.1310 | 0.1303 | 0.1308 | 0.1891 | 0.1880 | 0.1892 |
| | 10 | 0.0649 | 0.0645 | 0.0646 | 0.1433 | 0.1430 | 0.1439 |
| | 50 | 0.0170 | 0.0170 | 0.0169 | 0.1111 | 0.1107 | 0.1114 |

Estimated standard errors of intraclass correlations were from the Osborne and Paterson (1952) approximation based on 10⁵ replicated samples. Standard errors of mean estimates were <0.0002 (*h*² = 0.04) and <0.002 (*h*² = 0.96).

^a Heritability (four times the intraclass correlation).

^b Number of sires.

^c Number of progeny per sire.

^d Four times the empirical standard error of the estimated intraclass correlation.

heritability was 1.2619, whereas the predictions from Osborne and Paterson (1952) and Fisher (1921) were 0.2687 and 0.5183, respectively.

The average estimated standard errors of the heritability estimates are close to the observed standard error from simulation. Clearly, the approximation of Osborne and Paterson (1952), *i.e.*, the substitution of the estimated heritabilities into Equation 1, works very well, and, therefore, using Fisher's or Kempthorne's formula will underestimate the true standard error by factors of $[(s - 1)/s]^{1/2}$ and $[(s - 1)/(s + 1)]^{1/2}$. It appears that the average estimated standard error is closer to the observed standard error than the prediction using the population values of the heritability.

Sampling variances for genetic correlations: Results are presented in Table 2. Clearly for small *n*, the empirical standard error of *r_g* is usually larger than that predicted under the unconstrained (least-squares) model. For example, for *s* = 100 and *n* = 10, the empirical standard error is 0.833 for *h*² = 0.10 and *r_g* = 0.0, whereas the predicted value is 0.509. When the parameters are forced in the parameter space, the maximum empirical standard error is 1.0, when half of the time an estimate of +1 is obtained, and half of the time an estimate of -1. For small *s* and *n*, the empirical standard error can then be smaller than that predicted. For example, for *s* = 100 and *n* = 2, the empirical standard error from REML was 0.897 for *h*² = 0.10 and *r_g* = 0, whereas the predicted value was 2.84. For large *n* (>10), the equations perform well. Substituting the estimated parameters into expressions 8 and 9 is almost always worse, because of the real possibility of obtaining very small

estimates of the heritabilities. Only for large designs are the average estimated and predicted standard errors similar. For powerful designs, *i.e.*, for those designs with a small probability of obtaining least-squares estimates that are out of bounds, the average estimated sampling variance appears to be closer to the observed sampling variance than the predicted values (Table 2).

Relationship between heritability estimates and sampling variance of *r_g*: A more detailed investigation into the relationship between estimated heritabilities, estimated genetic correlation coefficients, and the sampling variance of the genetic correlation estimate was performed for *s* = 100, *n* = 1000, *h*² = 0.50 (both traits), and *r_w* = *r_g* = 0.75. One million replicated populations were simulated, and both the observed standard error and the estimated standard error were summarized as a function of the geometric mean of the heritability estimates (*i.e.*, $\{\hat{h}_1 \times \hat{h}_2\}$). Simulation results are displayed in Figure 1. The graph also includes a plot of the predicted standard error, assuming that the values of the heritabilities on the *x*-axis are the population values. Since for a powerful design with many progeny per sire the predicted sampling variance of the genetic correlation coefficient does not depend on the heritabilities (Equations 14 and 16), the corresponding line in Figure 1 appears to be horizontal. The prediction of the standard error using population values, *i.e.*, *h*₂ = 0.50 and *r_g* = *r_w* = 0.75, is 0.0443 for this design. The observed standard error of genetic correlation coefficients over all samples, hence also over all possible values of estimated heritabilities, was 0.0443, and the correlation between the estimated genetic correlation and the geomet-

TABLE 2
Empirical (*O*), estimated (*E*), and predicted (*P*) standard errors of estimated genetic correlation coefficients

| <i>s</i> ^a | <i>n</i> ^b | Least squares | | | | | REML | | | |
|-----------------------|-----------------------|------------------------------|--------------------------|-----------------------|------------------|------------------|--------------------------|------------------|------------------|------------------|
| | | $h_1^2 = h_2^2$ ^c | $r_g = r_p$ ^d | p_{LS} ^e | $O(\sigma(r_g))$ | $E(\sigma(r_g))$ | p_{PREML} ^f | $O(\sigma(r_g))$ | $E(\sigma(r_g))$ | $P(\sigma(r_g))$ |
| 100 | 2 | 0.10 | 0.00 | 36 | 2.625 | * | 20 | 0.897 | * | 2.84 |
| | | 0.10 | 0.50 | 39 | 1.846 | * | 20 | 0.834 | * | 2.13 |
| | | 0.25 | 0.00 | 53 | 2.124 | * | 31 | 0.823 | * | 1.14 |
| | | 0.25 | 0.50 | 56 | 1.526 | * | 31 | 0.768 | * | 0.853 |
| | 10 | 0.10 | 0.00 | 85 | 0.833 | 1.500 | 74 | 0.622 | 1.512 | 0.509 |
| | | 0.10 | 0.50 | 85 | 0.632 | 1.144 | 73 | 0.550 | 1.259 | 0.382 |
| | | 0.25 | 0.00 | 100 | 0.296 | 0.300 | 100 | 0.296 | 0.300 | 0.256 |
| | | 0.25 | 0.50 | 100 | 0.238 | 0.236 | 100 | 0.238 | 0.236 | 0.192 |
| 500 | 2 | 0.10 | 0.00 | 50 | 1.812 | 9.411 | 35 | 0.838 | 7.890 | 1.27 |
| | | 0.10 | 0.50 | 53 | 1.256 | 6.761 | 35 | 0.767 | 6.783 | 0.950 |
| | | 0.25 | 0.00 | 84 | 1.038 | 3.821 | 73 | 0.628 | 3.540 | 0.507 |
| | | 0.25 | 0.50 | 85 | 0.776 | 2.647 | 73 | 0.569 | 3.086 | 0.380 |
| | 10 | 0.10 | 0.00 | 100 | 0.259 | 0.259 | 100 | 0.259 | 0.259 | 0.227 |
| | | 0.10 | 0.50 | 100 | 0.207 | 0.202 | 100 | 0.207 | 0.202 | 0.170 |
| | | 0.25 | 0.00 | 100 | 0.116 | 0.116 | 100 | 0.116 | 0.116 | 0.114 |
| | | 0.25 | 0.50 | 100 | 0.088 | 0.087 | 100 | 0.088 | 0.087 | 0.086 |
| 1000 | 2 | 0.10 | 0.00 | 61 | 1.460 | 5.576 | 46 | 0.782 | 4.823 | 0.895 |
| | | 0.10 | 0.50 | 63 | 1.038 | 4.060 | 46 | 0.709 | 4.122 | 0.671 |
| | | 0.25 | 0.00 | 95 | 0.619 | 1.355 | 90 | 0.473 | 1.255 | 0.359 |
| | | 0.25 | 0.50 | 95 | 0.499 | 0.997 | 90 | 0.418 | 1.010 | 0.269 |
| | 10 | 0.10 | 0.00 | 100 | 0.170 | 0.168 | 0.00 | 100 | 0.170 | 0.168 |
| | | 0.10 | 0.50 | 100 | 0.130 | 0.128 | 100 | 0.130 | 0.128 | 0.120 |
| | | 0.25 | 0.00 | 100 | 0.082 | 0.081 | 100 | 0.082 | 0.081 | 0.081 |
| | | 0.25 | 0.50 | 100 | 0.061 | 0.061 | 100 | 0.061 | 0.061 | 0.061 |

Empirical and estimated standard errors of genetic correlations were estimated from 10^5 replicate populations. Predicted standard errors were calculated using equations from Tallis (1959) using the population values for the heritabilities and correlation coefficients. An asterisk (*) indicates that the estimate was >10.0 .

^a Number of sires.

^b Number of progeny per sire.

^c Heritability for both traits.

^d Genetic and phenotypic correlation coefficient

^e Proportion ($\times 100$) of replicates for which estimates of both sire variances were positive, and the geometric mean of the estimated heritabilities >0.01 .

^f Proportion ($\times 100$) of replicates for which the estimated genetic covariance matrix was positive-definite.

ric mean of the heritability estimates was 0.59 (results not shown).

DISCUSSION AND CONCLUSIONS

Population or estimated values: It is clear that all the expressions for the sampling variance of intraclass correlations or genetic correlation coefficients were essentially derived using a first-order Taylor series about the true population values. Hence, these are the values that should be used to study, for example, the power of various experimental designs, because they give the best prediction of the sampling variance.

Sampling variance of intraclass correlation: The prediction of the sampling variance of the heritability estimate based upon population values is accurate for small heritabilities and/or a large number of sires. However, for a small number of sires and a large heritability, the

prediction is very poor. The reason for this is that the distribution of the heritability estimate becomes very skewed in these cases, and the Taylor series in Equation 1 ignores higher-order terms by implicitly assuming that both numerator and denominator in the series are normally distributed. In fact, they are distributed proportionally to χ^2 distributions, which are known to be highly skewed for a small number of degrees of freedom [the coefficient of skewness of a central χ^2 distribution with k degrees of freedom is $2^{3/2}/k^{1/2}$ (Lancaster 1969, p. 20)]. Higher-order Taylor series converged only slowly (results not shown) and do not improve the prediction of the sampling variance substantially. For the extreme case of sires with many progeny it was shown (Equation 2) that the predicted standard error of the estimated heritability is proportional to $t(1 - t)$. However, the observed standard error in this case is a function of t , since $(B/n\sigma_p^2)$ is the only term that varies in the Taylor

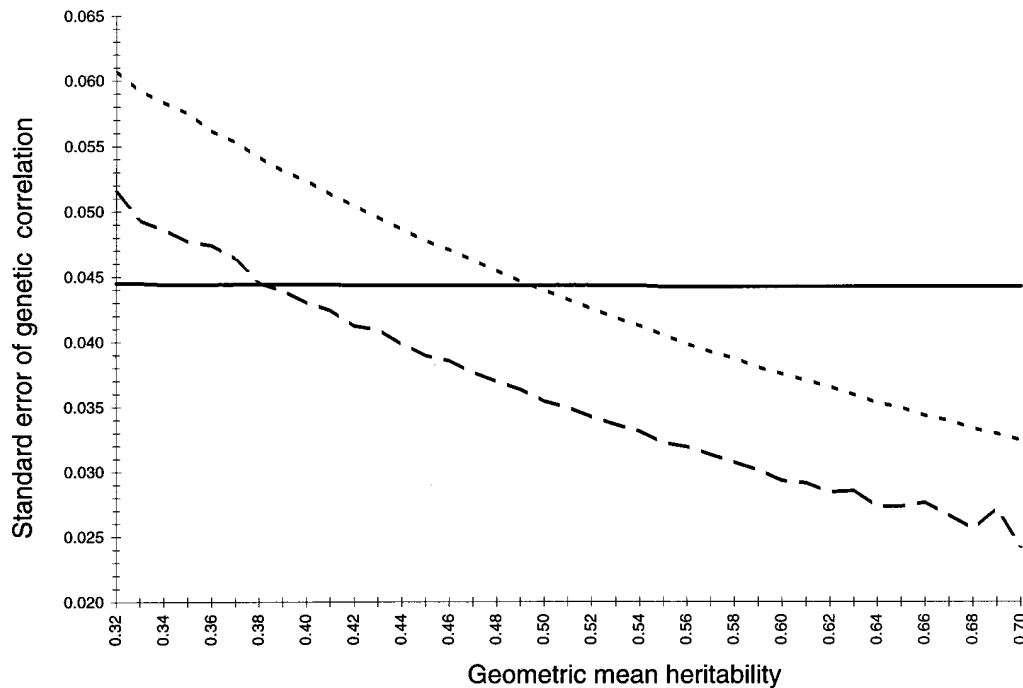


Figure 1.—Standard error of the genetic correlation coefficient against the geometric mean of the estimated heritabilities, for a design of $s = 100$ and $n = 1000$. Population values were 0.50 for the two heritabilities, and 0.75 for the phenotypic and genetic correlation coefficients. Simulated results based upon 10^6 samples. —, predicted values using Equation 14, assuming that the values on the x-axis are the population values; - - -, observed standard error; ···, estimated standard error, using the estimated values of the heritabilities and correlation coefficients.

series. Hence, the Osborne and Paterson (1952) approximation is biased downward for large heritabilities and a small number of sires. Although results were not shown, the estimate of the heritability is also biased downward in these cases. However, heritabilities of quantitative traits are not often >0.5 , and estimates are usually based on a reasonable number of families, so that this problem of an underprediction of the sampling variance is unlikely to occur.

Estimating the sampling variance of intraclass correlations by substituting the estimates of the heritabilities into the standard expressions (Osborne and Paterson 1952) works remarkably well, in that the average estimated standard error is very close to that predicted. For a large value of the heritability ($h^2 = 0.96$) and a small number of sires ($s = 2$ or $s = 4$), the average estimated standard error is smaller than the predicted standard error and appears to be close to the observed standard error (Table 1). However, further simulations using more extreme values of t showed that this is not a general observation. For example, for $t > 0.5$ (corresponding to "heritability" >2.0) the prediction using population values is closer to the observed sampling variance than the average estimated sampling variation (results not shown). Part of the reason why the average estimated standard error of the heritability is closer to the observed sampling variance than the prediction based upon t for heritability values in the normal range is that it takes account of the bias in the heritability estimate. Heritability estimates are biased downward, in particular for large values of t and small values of s . For example, for $h^2 = 0.96$, $s = 4$, and $n = 1000$, the average estimate for the heritability from simulation is 0.86 (results not in

tables). When this value is used for the standard prediction equation (Equation 1), the predicted standard error of the heritability is 0.5532, which is closer to the observed standard error (0.5169, Table 1) than that predicted from the population value of the heritability (0.5978, Table 1).

Sampling variances of genetic correlations: The expressions for $\text{var}(r_g)$ from Reeve, Tallis, and Robertson perform poorly using small population sizes and small heritabilities when the known population values are used to predict the sampling variance of the genetic correlation coefficient. This is because the estimates of the genetic correlation coefficient can become very large (positive or negative) when using least-squares methods. Using REML, the equations perform much better, although the empirical standard errors are generally larger than those predicted in Table 2. Substituting the estimates of the heritabilities and genetic correlations into Equation 15 can result in a very large estimate of the sampling variance of the genetic correlation when there is a real chance that the estimates of the heritabilities approach zero (the numerator of Equation 15 approaches $2/[n(s-1)(n-1)]$ for both intraclass correlations approaching zero, whereas the denominator goes to zero).

Koots and Gibson (1996) showed in one of their simulation studies that the empirical sampling variance of the genetic correlation coefficient depended on the estimates of the two heritabilities and concluded that, therefore, the estimates of the heritabilities should be used in, for example, Equation 15, even if the population values were known. Further simulation results using the same population values (K. Koots and J. Gibson,

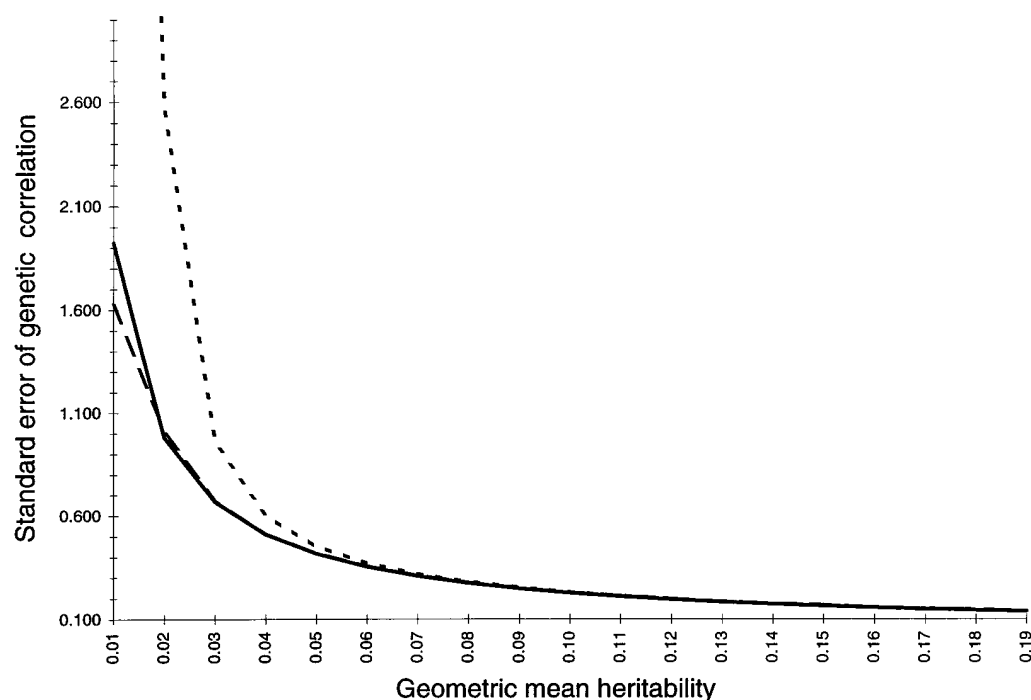


Figure 2.—Standard error of the genetic correlation coefficient against the geometric mean of the estimated heritabilities, for a design of $s = 500$ and $n = 10$. Population values were 0.10 for the two heritabilities, and 0.0 for the phenotypic and genetic correlation coefficients. —, predicted values using Equation 14, assuming that the values on the x -axis are the population values; ---, observed standard error; ···, estimated standard error, using the estimated values of the heritabilities and correlation coefficients.

personal communication) showed very clearly that the sampling variance of the genetic correlation coefficient, conditional on the values of the estimated heritabilities, is accurately estimated by substituting the estimated (and not the true) heritabilities into Equation 15, in the case of a parent-offspring design, heritabilities of 0.10, and environmental and genetic correlation coefficients of zero. This appears to contradict the simulation results in Table 2, which show that the average sampling variance is poorly estimated using estimated parameters. However, the results in Table 2 indicate that the estimation is poor only when the least-squares estimates are likely to be outside the parameter space. In the cases where the least-squares and REML estimates are identical, both the prediction and average estimated sampling variance are close to the observed one.

The relationship between estimated heritabilities and the empirical sampling variance of the genetic correlation was explored in Figure 1 for a powerful design. From these results we may conclude that (1) the predicted sampling variance accurately predicts the average observed sampling variance (0.0443), but not the observed sampling variance for given values of the achieved estimates of the heritabilities; (2) for a given value of the estimates of the heritabilities, the estimated sampling variance follows a very similar pattern to that of the observed sampling variance (as in Koots and Gibson 1996); (3) for a given value of the estimates of the heritabilities, the estimated sampling variance is larger than the observed sampling variance; and (4) when the estimated heritabilities are close to the true values, the predicted and estimated sampling variances coincide. For other designs, a similar pattern was ob-

served. When the experiment was large and the population values of the correlation coefficients zero, the observed and estimated sampling variances, as a function of the geometric mean of the estimated heritabilities, were very similar. These additional results confirm the results of Koots and Gibson (1996) and reinforce their recommendation that the value estimated heritabilities should be used in calculating the sampling variance of the genetic correlation coefficient, even in the rare cases when the population values of the heritabilities are known. From Table 2 it appears that the average estimated sampling variance of the genetic correlation coefficient is closer to the observed sampling variance than the prediction using population parameters. For one design, $s = 500$, $n = 10$, $h^2 = 0.10$, $r_g = r_w = 0$, this was further explored using the results from 10^6 replicated samples. Figure 2 shows that for $\hat{h}_1\hat{h}_2 > 0.07$, the observed, estimated, and predicted sampling variance are virtually identical. For each of the geometric mean classes, the mean estimate of the genetic correlation was zero (results not shown), so that for a particular value of the estimated geometric mean of the heritabilities, the sampling variance of the genetic correlation coefficients reflects sampling from a population with the true values of the heritabilities equal to those estimated. This is in contrast with the previous example, in which the correlation between the estimated genetic correlation and geometric mean of the heritabilities was positive (+0.59). For smaller values of the estimated heritabilities, Figure 2 shows that the estimated sampling variance is much larger than either the predicted or observed sampling variance. The estimated sampling variance of the genetic correlation coefficient for

$\hat{h}_1\hat{h}_2 = 0.01$ was 9.9. It is not clear from these results why the average estimated sampling variance is closer to the observed sampling variance than the prediction using population values.

van Vleck and Henderson (1961) investigated the behavior of the expression derived by Reeve (1955) for the parent-offspring regression scenario by simulation. They came to the same conclusion as Koots and Gibson (1996); *i.e.*, in the case of a single progeny and one parent, more than 1000 pairs were needed before Reeve's expression was reasonably accurate.

Bias in estimate of sampling variance: It is usually assumed that by substituting the estimates of population parameters in the prediction equations for the sampling variance of the heritability or genetic correlation, unbiased estimates of those sampling variances are obtained. However, this is not generally the case for small sample sizes. Furthermore, when comparing the average sampling variation from simulation, it matters whether results are expressed in the average standard error (as in this study) or in the average sampling variance. This is best illustrated using the example of a half-sib design with large n and a small population value of t . Then

$$\text{var}(\hat{t}) = 2t^2/(s-1) \quad \text{and} \quad \sigma(\hat{t}) \approx t\sqrt{2}/(s-1),$$

with corresponding estimates,

$$\hat{\text{var}}(\hat{t}) \approx 2\hat{t}^2/(s-1) \quad \text{and} \quad \hat{\sigma}(\hat{t}) \approx \hat{t}\sqrt{2}/(s-1).$$

If expectations are taken over these estimates, then

$$\begin{aligned} E[\hat{\text{var}}(\hat{t})] &= E[2\hat{t}^2/(s-1)] \\ &= (2/(s-1))[\text{var}(\hat{t}) + E^2(\hat{t})] \\ &= (2/(s-1))[2t^2/(s-1) + t^2] \\ &= \text{var}(\hat{t})(s+1)/(s-1) \\ E[\hat{\sigma}(\hat{t})] &= E[\hat{t}]\sqrt{2}/(s-1) \\ &= t\sqrt{2}/(s-1) = \sqrt{\text{var}(\hat{t})}. \end{aligned}$$

Hence, for small t and large n , the estimate of the sampling variance can be severely biased upward for a small number of sires, whereas the estimate of the standard error is unbiased. It follows that the adjustment of the degrees of freedom suggested by Kempthorne (1957, Equation 8) gives an unbiased estimate of the sampling variance for small t , but a severely biased estimate of the standard error. In practice one should therefore be cautious when comparing or using estimated sampling variances from different experiments. In particular, combining heritability estimates from different-sized experiments by weighting the estimates proportionally to the inverse of the estimated sampling variance should be avoided because of an induced positive correlation between the estimate of the heritability and the estimate of its sampling variance, following Equations 1 and 7 (W. G. Hill, personal communication). This induced correlation causes the combined heritability estimate to

be biased downward, because of the negative correlation between the estimate of the heritability and its weight (the inverse of the sampling variance). Furthermore, a downward-biased overall heritability estimate would be obtained because the heritability estimate from smaller experiments tends to be biased downward, and the corresponding estimated sampling variance would be too small, giving too much weight to the smaller experiments. A joint analysis of data, or an iterative procedure, in which the estimated sampling variance for each experiment is recalculated from the pooled heritability estimate (W. G. Hill, personal communication) is to be preferred. However, a further complication arises because of a bias in the estimated sampling variance if the standard prediction equations are used. This bias may be upward (as shown above) or downward depending on the population parameters. To avoid strong biases in the pooled heritability estimates, a single data analysis should be carried out.

Conclusion: For the design of experimental populations to estimate genetic parameters, the prediction of the sampling variance of heritabilities using Osborne and Paterson (1952) is accurate, unless the population heritability is large and the number of family groups is very small. For analysis of data, the estimate of the standard error of the heritability obtained by substituting the estimated heritability for the true value in the standard prediction formulas is almost unbiased for the range of heritabilities and sample sizes likely to be encountered in practice.

For small experiments, estimates of heritability are biased downward, and estimates of sampling variances are generally not unbiased. Combining results from different experiments by weighting the heritability estimates by the inverse of their estimates-sampling variances may result in a severely biased heritability estimate, because the smaller experiments tend to have estimates that are too low, and too much weight is given to these estimates if their sampling variances are biased downward too. A joint analysis of all data is to be preferred.

The predicted sampling variance of the genetic correlation using Reeve (1955) and Tallis (1959) are accurate only if the population heritabilities are not close to zero and if the number of families is large. Even if the population heritabilities are known, the estimated heritabilities should be used in the estimation of the sampling variance of the genetic correlation coefficient.

I thank Naomi Wray and Bill Hill for helpful comments and Ken Koots and John Gibson for many constructive discussions and the sharing of additional simulation results.

LITERATURE CITED

- Calvin, J. A., 1993 REML estimation in unbalanced multivariate variance components models using an EM algorithm. *Biometrics* 49: 691-701.

- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Longman, Harlow, UK.
- Fisher, R. A., 1921 On the probable error of a coefficient of correlation deduced from a small sample. *Metron* **1**: 3–32.
- Hayes, J. F., and W. G. Hill, 1981 Modification of estimates of parameters in the construction of genetic selection indices ("bending"). *Biometrics* **34**: 429–439.
- Kempthorne, O., 1957 *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.
- Koots, K. R., and J. P. Gibson, 1996 Realized sampling variances of estimates of genetic parameters and the difference between genetic and phenotypic correlations. *Genetics* **143**: 1409–1416.
- Lancaster, H. O., 1969 *The Chi-Squared Distribution*. John Wiley & Sons, New York.
- Meyer, K., and W. G. Hill, 1991 Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. *J. Anim. Breed. Genet.* **109**: 264–280.
- Osborne, R., and W. S. B. Paterson, 1952 On the sampling variance of heritability estimates derived from variance analyses. *Proc. R. Soc. Edinb. Sect. B* **64**: 456–461.
- Patterson, H. D., and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545–554.
- Reeve, E. C. R., 1955 The variance of the genetic correlation coefficient. *Biometrics* **11**: 357–374.
- Robertson, A., 1959a The sampling variance of the genetic correlation coefficient. *Biometrics* **15**: 469–485.
- Robertson, A., 1959b Experimental design in the evaluation of genetic parameters. *Biometrics* **15**: 219–226.
- Robertson, A., and I. M. Lerner, 1949 The heritability of all-or-none traits: viability of poultry. *Genetics* **34**: 395–411.
- Tallis, G. M., 1959 Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Aust. J. Stat.* **1**: 35–43.
- Taylor, S. C., 1976 Multibreed designs. 1. Variation between breeds. *Anim. Prod.* **23**: 133–144.
- van Vleck, L. D., and C. R. Henderson, 1961 Empirical sampling estimates of genetic correlations. *Biometrics* **17**: 359–371.
- Visscher, P. M., 1995 Bias in genetic R^2 from halfsib designs. *Genet. Sel. Evol.* **27**: 335–345.
- Zerbe, G. O., and D. E. Goldgar, 1980 Comparison of intraclass correlation coefficients with the ratio of two independent F-statistics. *Commun. Stat. Theor. Meth.* **A9**: 1641–1655.

Communicating editor: R. G. Shaw