

Possible Biases in Heritability Estimates from Intraclass Correlation

R.W. Ponzoni and J.W. James

School of Wool and Pastoral Sciences, University of New South Wales, Kensington (Australia)

Summary. There is an inherent bias in intraclass correlations since the expectation of a ratio does not equal the ratio of expectations. A simple accurate approximation for this bias is derived, and it is found that the inherent bias is usually negligible. Selection of sires is known to bias half-sib heritability estimates, and appropriate formulae are given and discussed.

Key words: Bias – Heritability – Intraclass correlation – Selection

Introduction

Heritability is often estimated from the intraclass correlation between half-sibs or full-sibs. Half-sib estimates are largely free of bias from non-additive components of genetic variance and common environmental effects but in general full-sib estimates are not. However, not much attention has been paid to the fact that the intraclass correlation is inherently biased, since it is a ratio of variables. The ratio of the expectations of these variables is the true heritability but the expectation of the ratio is not the same as the ratio of the expectations. Pederson (1971) has given some numerical examples and Ginsburg (1973) has provided a general solution. Ginsburg's exact formula is very complicated and in this paper a simple but accurate approximation is presented. A better-known source of bias is that due to selection of sires in half-sib heritability estimates. This has been discussed in various ways by Reeve (1953), Morley (1955) and Rønningen (1972). The last-named gave a wrong formula and in this paper we give the correct one. The effect of selection for one trait on bias in heritability estimates for another trait is also studied.

Inherent Bias

The experimental design we consider is shown in Table 1, where the various parameters are defined. The true intraclass correlation t is estimated by

$$\hat{t} = \frac{X}{Y}$$

where $X = V_s - V_w$, and $Y = V_s + (n-1)V_w$

$$\text{Now } t = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2} = \frac{E(X)}{E(Y)} = \frac{\bar{X}}{\bar{Y}}$$

Table 1. Analysis of variance of sib data

Source of variation	d.f.	M.S.	Expected value
Between sib groups	s-1	V_s	$\sigma_w^2 + n\sigma_s^2$
Within sib groups	s(n-1)	V_w	σ_w^2

s = number of sib groups

n = number of sibs per group

V_s and V_w are observed mean squares

σ_s^2 and σ_w^2 are variance components between and within sib groups

Letting dX and dY denote deviations of X and Y from their means we have

$$\hat{t} = t(1 + dX/\bar{X}) / (1 + dY/\bar{Y})$$

and therefore

$$\frac{\hat{t}}{t} = 1 + \frac{dX}{\bar{X}} - \frac{dY}{\bar{Y}} - \frac{dXdY}{\bar{X}\bar{Y}} + \frac{dY^2}{\bar{Y}^2} + \dots$$

Taking expectations and ignoring third and higher order terms,

$$E\left(\frac{\hat{t}}{t}\right) \approx 1 - \frac{\text{cov}(XY)}{\bar{X}\bar{Y}} + \frac{\text{var}(Y)}{\bar{Y}^2}$$

Now $\text{cov}(XY) = 2V_s^2/(s-1) - 2V_w^2/s$

and $\text{var}(Y) = 2V_s^2/(s-1) + 2(n-1)V_w^2/s$

and on substituting and simplifying we find

$$E(\hat{t}-t) \approx \frac{-2(1-t)\left(t + \frac{1-t}{n}\right)\left(t + \frac{1-t}{sn}\right)}{s-1} \quad (1)$$

Thus there is a negative bias in the estimate which depends on the correlation itself and on the family structure.

The ratio of the absolute value of the bias to the standard error of the estimate is

$$\frac{|\text{Bias}|}{\text{S.E.}} = \sqrt{2} \left(t + \frac{1-t}{T}\right) \sqrt{\frac{T-s}{(T-1)(s-1)}} \quad (2)$$

where $T = sn$ is the total number of progeny. Provided s is a good deal smaller than T the order of magnitude of this ratio is t/\sqrt{s} which means that unless s is very small, the bias is negligible compared with the standard error. For example, if Robertson's (1959) optimum design is used, where $s = Tt$, then for small t and large T the ratio is about $\sqrt{2t/T}$, and the bias could be safely ignored.

Numerical calculations showed that the approximation (1) gave excellent agreement with Ginsburg's (1973) exact results. The bias is larger for high intraclass correlations and small numbers of families but will not be of practical importance in any experiment where an accurate heritability estimate is obtained. It may sometimes be worth correcting an estimate based on a small body of data by using equation (1).

Bias due to Sire Selection

If a half-sib estimate of heritability is derived from an experiment in which sires have been selected for a trait X so that after selection their phenotypic variance is $\sigma_x^2(1-K)$, the estimate of heritability of X will be biased. When selection is by truncation of the distribution of X at x standard deviations from the mean to give a standardised selection differential i , it is well known that $K = i(i-x)$. If Y denotes the expected progeny mean of the trait for a sire, then for unselected sires $\sigma_y^2 = \frac{1}{4}h_x^2 \sigma_x^2$ and the correlation between X and Y is h_x , where h_x^2 is the heritability of the trait. Cochran (1951) showed that when X is selected, the proportional reduction in variance of a correlated trait is the product of the proportional reduction in variance of the selected trait and the squared correlation between the two traits. Thus the proportional reduction in variance of Y is Kh_x^2 , so the between sire component of variance

would become $\frac{1}{4}h_x^2 \sigma_x^2 (1-Kh_x^2)$. The variance within families would be unaffected at $\sigma_x^2 (1-\frac{1}{4}h_x^2)$ and thus the total phenotypic variance among progeny would be $\sigma_x^2 (1-\frac{1}{4}Kh_x^4)$. So the heritability estimated from selected sires, h_{xs}^2 , relative to the true heritability is

$$\frac{h_{xs}^2}{h_x^2} = \frac{1 - Kh_x^2}{1 - \frac{1}{4}Kh_x^4} = \frac{1 - Q}{1 - \frac{1}{4}h_x^2 Q} \quad (3)$$

where $Q = Kh_x^2$.

If Z denotes the expected progeny mean of a second trait W , with unselected sires $\sigma_z^2 = \frac{1}{4}h_w^2 \sigma_w^2$ and the correlation between X and Z is $r_g h_x$, where h_w^2 is the heritability of W and r_g is the genetic correlation between X and W . Therefore the between sire component of variance for W when sires are selected on X will be $\frac{1}{4}h_w^2 \sigma_w^2 (1 - Kr_g^2 h_x^2)$. The variance within families will be $\sigma_w^2 (1-\frac{1}{4}h_w^2)$ and thus the heritability for W estimated from sires selected on X , h_{ws}^2 , relative to the true heritability, is

$$\frac{h_{ws}^2}{h_w^2} = \frac{1 - Kr_g^2 h_x^2}{1 - \frac{1}{4}h_w^2 Kr_g^2 h_x^2} = \frac{1 - Q}{1 - \frac{1}{4}h_w^2 Q} \quad (4)$$

where $Q = Kr_g^2 h_x^2$.

Q is the proportional reduction in the variance of expected progeny means and both (3) and (4) have the same form when Q is so defined. Thus the bias in a correlated trait is about r_g^2 times that in the selected trait, and unless the traits are highly correlated, selection on X would not severely bias the apparent heritability of W . For example, if $h_x^2 = h_w^2 = 0.4$, and $r_g = 0.5$, when $K = 0.86$ (5% selected), $h_{xs}^2 = 0.68h_x^2$ while $h_{ws}^2 = 0.92h_w^2$.

If r_p is the phenotypic correlation between X and W then the phenotypic variance of W among sires selected on X is $\sigma_w^2 (1-Kr_p^2)$. However, this observed reduction in variance of W cannot be used to calculate the bias in estimation of h_w^2 . If it were assumed that this could be done the apparent ratio of heritabilities would be

$$(1 - Kr_p^2 h_w^2) / (1 - \frac{1}{4} Kr_p^2 h_w^4).$$

This ratio would be correct only if $r_g h_x = r_p h_w$. Thus adjustment for sire selection can only be done if the trait on which selection is based is definitely known.

Discussion

It is clear that the inherent bias is small enough to be of no importance in most circumstances. However, on occasions, heritability estimates are obtained for measurements which are very expensive or time-consuming and for this reason may be based on small numbers. For example, Buckland (1975) presented heritability estimates

based on 8 families each of 3 full-sibs. For a heritability of 0.5 this would produce a bias of -0.06 , which, though of some concern, is hardly large enough to be very important when the accuracy is so low, the bias being only one eighth of the standard error. Such a bias might be of concern if data from several such experiments were pooled by averaging heritability estimates, rather than by pooling sums of squares. Equation (1) could then be used to adjust individual estimates if heritability estimates were to be averaged because mean squares were not reported.

The formula given by Rønningen (1972) for bias due to sire selection is wrong. Rønningen's error stems from the fact that equation (11) of Cochran (1951) gives the correlation between genetic value and phenotypic value among selected individuals. But what we actually estimate is the correlation among progeny from those selected parents. Among these progeny there will be further genetic variance due to segregation.

Figure 1 shows the ratio of heritability estimated from selected sires to true heritability, for a range of heritability values and proportions selected. The figure shows clearly how severe the bias may be when sires are selected intensely for a highly heritable character. Under these conditions equation (3) could be used to adjust the heri-

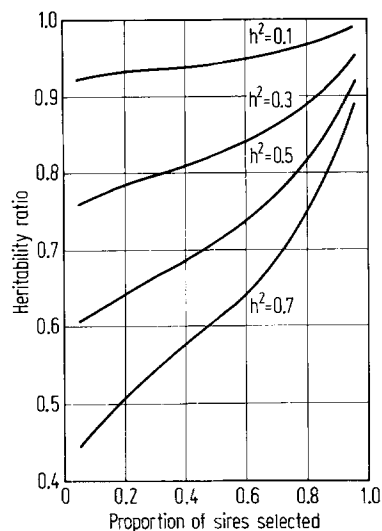


Fig. 1. The ratio of heritability estimated from progeny of selected sires to true heritability (h_{sx}^2/h_x^2) plotted against proportion of sires selected for four different true heritabilities

tability estimate from selected sires to the value it would have been when sires are randomly chosen. If sires have been selected on a correlated trait, equation (4) could be used, if an estimate of the genetic correlation were available. Since accurate estimates of genetic correlations are unlikely to be available when it is still necessary to estimate heritability, this adjustment method is likely to be less generally useful. It is important that the difference in adjustment between the character selected and other characters should be recognised.

Acknowledgment

This work was done while R.P. held a F.A.O./U.N.D.P. Fellowship.

Literature

- Buckland, R.B.: Genetic influences on the activity of some enzymes in fresh and stored chicken semen and their relationships to fertility. In: R. Bogart (Ed.), *Genetics Lectures*, Vol. 4, 39-67. Corvallis: Oregon State U.P. 1975
- Cochran, W.G.: Improvement by means of selection. Proc. 2nd. Berkeley. Symp. Math. Statist. Probab. 1950, 449-470 (1951)
- Ginsburg, E.H.: On the planning of the experiment on estimation of intraclass correlation. *Biom. Z.* 15, 47-52 (1973)
- Morley, F.H.W.: Selection for economic characters in Australian Merino sheep. V. Further estimates of phenotypic and genetic parameters. *Aust. J. agric. Res.* 6, 77-90 (1955)
- Pederson, D.G.: The estimation of heritability and degree of dominance from a diallel cross. *Heredity* 27, 247-264 (1971)
- Reeve, E.C.R.: Studies in quantitative inheritance. III. Heritability and genetic correlation in progeny tests using different mating systems. *J. Genet.* 51, 520-542 (1953)
- Rønningen, K.: The effect of selection of progeny performance on the heritability estimated by half-sib correlation. *Acta Agric. Scand.* 22, 90-92 (1972)

Received May 8, 1978

Communicated by J.S.F. Barker

R.W. Ponzoni
Department of Agriculture and Fisheries
Box 1671, G.P.O.
Adelaide, S.A. 5001 (Australia)

J.W. James
School of Wool and Pastoral Sciences
University of New South Wales
Kensington, N.S.W. 2033 (Australia)